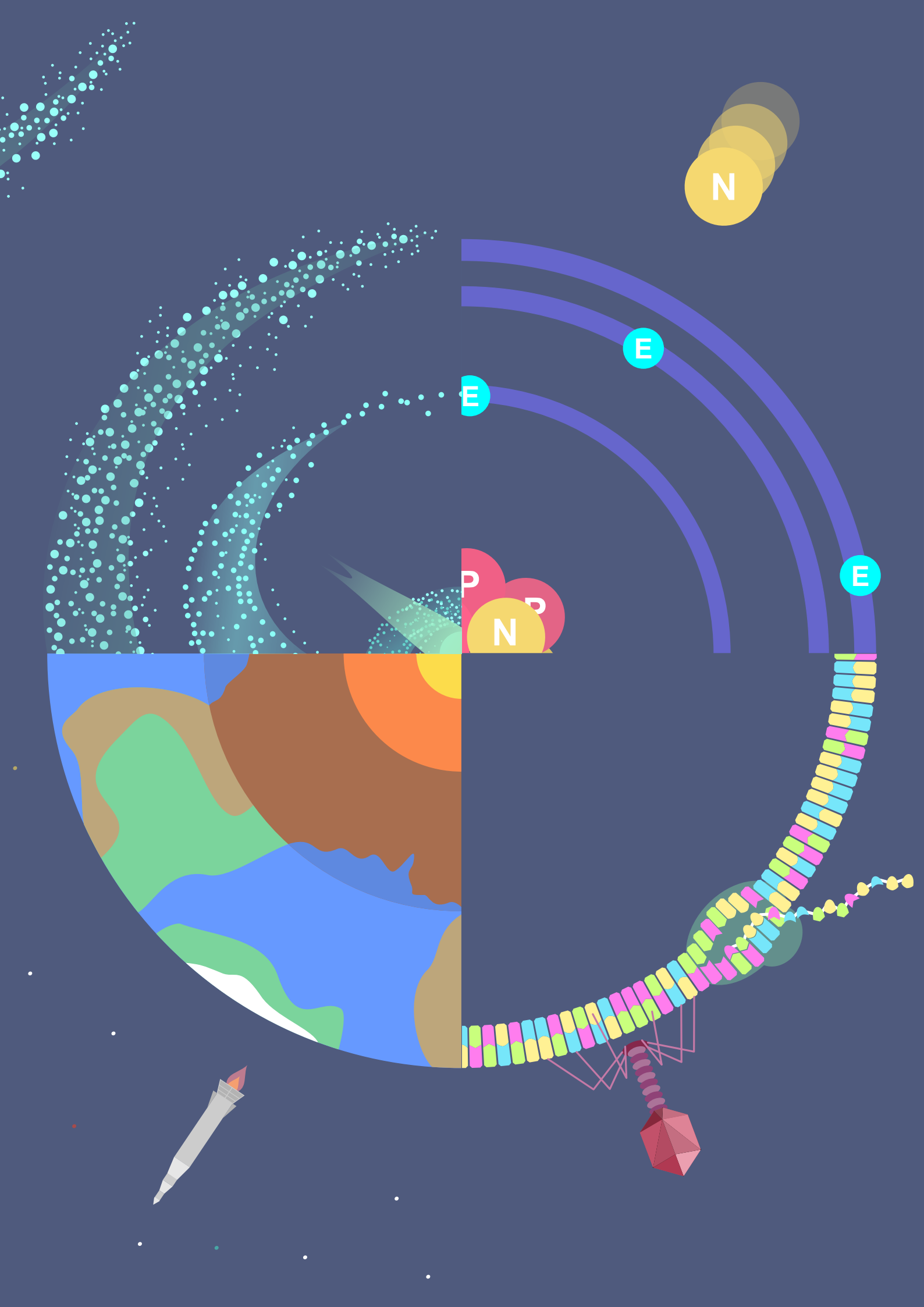
Evolutionary divergence of the insect disease-encoding *Serratia* plasmid pADAP

Lincoln University
Doctor of Philosophy Thesis

2020

Thomas L. Sitter


N
E
E
E
P
N

Evolutionary divergence of the insect disease-encoding *Serratia* plasmid pADAP

A thesis submitted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy

at Lincoln University
by Thomas Lesley Sitter

Lincoln University 2020

Abstract of a thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Evolutionary divergence of the insect disease-encoding *Serratia* plasmid pADAP

by

Thomas Lesley Sitter

The larval stage of the endemic New Zealand grass grub (*Costelytra giveni*) (Coleoptera: Scarabaeidae), causes estimated damage worth between 215 – 585M NZD annually to dairy and meat farms throughout New Zealand. To date, two diseases of grass-grub larvae, instigated by strains of *Serratia entomophila* and *S. proteamaculans*, have been identified and are being used as commercial biocontrol agents. As chemical pesticides are rapidly being banned due to hazardous side effect or soil contamination, together with documented increases of resistance to these pesticides, biological control approaches such as the *Serratia* based products are slowly taking their place.

The two main virulence determinants in these *Serratia* strains, an Anti-feeding prophage (Afp) and an ABC-toxin-complex (*Serratia entomophila* pathogenicity toxin complex), are encoded on the previously described 153-kb conjugative megaplasmid pADAP (Amber Disease Associated Plasmid). Variants of the pADAP have been identified containing only one of the two virulence determinants and homologs have been found in bacteria from other genera. With the goal of defining evolutionary points of divergence between plasmid variants, and to examine potential co-evolution between plasmid and host, plasmids from 76 *Serratia* isolates and three *Yersinia* isolates with varying disease phenologies were sequenced.

Phylogenetic analysis of the conserved pADAP plasmid "backbone", residing between a conserved point of demarcation found in this study and the end of a sex pili cluster (*spa1*), revealed clustering of all the *S. entomophila* plasmids. Within the predicted backbone region, several intergenic regions in the areas of replication and conjugation contained DNA inserts, one of which, positioned between TraG and TraC, demarcates chronic disease related plasmids from the hypervirulent and non-pathogenic plasmids. These inserted regions, together with other insertions, deletions and mutations identified in this project, are clear markers for evolutionary divergence between the plasmid variants.

In addition to virulence determinant variants, several novel gene clusters have been identified. Some of these include antimicrobials, putative accessory virulence determinants, toxin-antitoxin clusters and secretion systems, and two novel regions of unknown function. In this study, several of the novel regions were investigated to determine correlation between disease states in *C. giveni* and the occurrence of these genes, but none of the regions were found to influence amber disease in *C. giveni*, suggesting these strains might be associated with other hosts or provide the bacteria with other competitive advantages.

Of note is a bacteriophage present on a pADAP-type plasmid, that is also present in a non-pADAP type plasmid, both obtained from *S. proteamaculans* strains sequenced in this project. The non-pADAP phage carrying strain also has an orthologous Sep region. This phage and Sep carrying *S. proteamaculans* isolate was ineffective towards *C. giveni* larvae but had significant mortality response in *Pyronata festiva*, a beetle species morphologically similar to grass grub, but genetically very distinct. This is an interesting finding, as most Sep carrying strains without Afp have the opposite phenotype in that they are only bioactive in *C. giveni* and not *P. festiva*.

Numerous Sep variants were observed, including several variants residing on non-pADAP plasmids (i.e. without the conserved backbone). All but one of the Sep variants were co-located with the Sef fimbrial cluster, which was highly conserved compared to the Sep cluster. The one non-Sef encoding plasmid was associated with a *Yersinia frederiksenii* isolate and appeared to have no bioactivity in *C. giveni*. The co-location of the Sef cluster, the high conservation of the *sef* genes and the absence of plasmids lacking the Sef region while still being bioactive, leads to the assumption that fimbriae are an integral part to the functionality of Sep, through a yet to be determined mechanism.

The hypothesis of the study was that there would be a tightly regulated network of HGT mechanisms that made pADAP a clonal plasmid capturing device with actively interchangeable pathogenicity associated clusters. It was assumed virulence determinants can be lost in situations where the bacterium and its plasmid were not in contact with the insect host, or that the plasmid could be discarded in periods of low nutrients, such as the summer period in New Zealand where grass grub larvael populations are at their lowest and nutritional stress and heat stress are at their highest. However, the exact opposite of this hypothesis was observed.

The plasmids analyzed appear to be highly stable and loss of plasmid happens rarely even under the hashest conditions (extreme heat, extreme stress, nutritional depletion). Numerous analysed isolates with peculiar phenotypes thought to have a pADAP harboured non-pADAP megaplasmids of similar size. This study also showed that pADAP plasmids do not seem to be a burden to their host cell, but instead actually convey some fitness benefits. Instead of observing active horizontal gene transfer (HGT) of pathogenicity clusters, it was observed that the genetic diversity is much more correlated with speciation.

The overall conclusion of this study is that perturbation of the bacterium system is more likely caused by genetic diversity than an tightly regulated active HGT mechanisms.

Keywords: Insecticidal activity, plasmids, evolutionary divergence, horizontal gene transfer, pan-genome, genome wide association study, grass grub, pADAP, Sep, Afp.

Acknowledgements

I would like to thank the following people for their contribution to this study;

The supervisors, Dr. Mark Hurst, Prof. Travis Glare, Prof. Peter Fineran, Prof. Murray Cox, Dr. Karen Armstrong and Dr. Paul Gardner. You all have guided me through this project, given me tips and tricks for experiments, advice on how to tackle hurdles in the project, point out mistakes or logical fallicies, helped me with my writing and presenting skills, but have also given me the freedom to let me take my project into my own hands and let me stear it in a direction that interested me.

Marion Schoof, Pushpanjali Bhardwaj, Amber Paulson and Renee Watson for their friendship and support throughout the project, as well as their help in troubleshooting, optimizing and brainstorming numerous experiments.

Amy Vaughan, for her collaboration and insights on the chromosomal side of isolates analysed in this study.

Richard Townsend for sharing his knowledge and expertise on the beetle species used in this study.

Amy Beattie, Mitchell Weston and Sandra Jones for help with lab protocols and equipment as well as support wit troubleshooting.

Dr. Marina Richena for her help with setting up the TEM experiments and troubleshooting image capturing problems.

Dr. Simon Jackson for his help with the flow cytometry analysis performed in this study.

Dr. Shikako van Kooten for her help explaining the proper method of analysing bioassay data.

BPRC for providing the funding to complete this work as well as opportunities to attend conferences to present my work.

AgResearch for hosting me, and providing me with the resources needed to carry out the work

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	Amino Acid
Afp	Anti-feeding prophage
ALA	5-aminolevulinic acid
ALA50	5-aminolevulinic acid 50 µg per ml
Ap400	Ampicillin 400 µg per ml
Ap^R	Ampicillin resistance
BLAST	Basic Local Alignment Search Tool
bp	Basepair
BRIG	BLAST Ring Image Generator
BT	*Bacillus thuringiensis*
CCS	Circular Consensus Read
CDS	Coding sequence
CLR	Continuous Long Read
Cm90	Chloramphenicol 90 µg per ml
Cm^R	Chloramphenicol resistance
DNA	Deoxyribonucleic acid
DUF	Domain of Unknown Function
EM	Extracellular matrix / milieu
gDNA	Genomic DNA
GE	Gel Electrophoresis
GFPmut3	Green Fluorescent Protein, mutant #3
GI	Genomic island
Gm15	Gentamicin 15 µg per ml
Gm^R	Gentamicin resistance
HGT	Horizontal Gene Transfer
HKY85	Hasegawa-Kishino-Yano 1985 nucleotide substitution model
IRs	Inverted repeat sequences
Kb	Kilobase
Km50	Kanamycin 50 µg per ml
Km^R	Kanamycin resistance
LB	Lysogeny-broth
LB-agar	Merck Miller Luria-Bertani base agar (Miller LB agar)
LB-broth	Invitrogen Lennox L Broth Base (LB Broth Base)
Mb	Megabase
MGE	Mobile Genetic Elements
MitC	Mitomycin C
ml	Millilitre
ML	Maximum Likelihood
mRNA	Messenger RNA
NaAc	Sodium Acetate
NCBI	National Centre for Biotechnology Information
ncRNA	Non-coding RNA
nr	Non-redundant protein database
o/n	Overnight
OM	Outer membrane
ORF	Open reading frame
pADAP	Amber disease associated plasmid
PAI	Pathogenicity associated islands
PCR	Polymerase chain reaction
PSK	Post-segregational killing
PVC	*Photorhabdus* virulence cassettes

RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNAP	RNA Polymerases
RUF	Region of unknown function
SACA	*Serratia* associated capturing apparatus
SDS-PAGE	Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis
Sea	*Serratia entomophila* annotation
Sef	*Serratia entomophila* fimbriae
Sep	*Serratia entomophila* pathogenicity
SGE	Selfish Genetic Elements
Sm^R	Streptomycin resistance
Sp160	Spectinomycin 160 µg per ml
Spp	*Serratia proteamaculans* pathogenicity
Sp^R	Spectinomycin resistance
TA	Toxin-Antitoxin
TC	Toxin complex
Tc30	Tetracycline 30 µg per ml
Tc^R	Tetracycline resistance
TEM	Transmission Electron Microscopy
tRNA	Transfer RNA
$tRNA^{fMet}$	Transfer RNA N-Formylmethionine
UA	Uranyl acetate
VGT	Vertical Gene Transfer
X-gal	5-bromo-4-chloro-3-indolyl--D-galactoside
X-gal50	5-bromo-4-chloro-3-indolyl--D-galactoside 50 µg per ml

Chapter 1 Historical context and general introduction

1.1 *Costelytra giveni*

The beetle, *Costelytra giveni* (1), also called Tutaeruru by the Māori and colloquially referred to as the grass grub, is an endemic insect to New Zealand. This species was described by Europeans in the years between 1839 to 1843, when the British royal navy set two warships, the H.M.S. Erebus and the H.M.S. Terror, out to the southern hemisphere to investigate terrestrial magnetism. The expedition was accompanied by scientists charged with documenting new marine and terrestrial organisms. One of the scientists, Adam White, was in charge of documenting invertebrates and in 1844 published his findings as part of a two volume book containing detailed descriptions and drawings of all new organisms found during the expeditions. One of these descriptions is believed to be the first taxonomic classification of the New Zealand grass grub, in that time it had been labeled *Rhistorogus zealandica* (2). In 1952 B.B. Given compared the description written by White, to 400 samples obtained from around New Zealand and found the specimens belonged to the genus *Costelytra*, not *Rhistorogus*, and the beetle was therefore renamed *Costelytra zealandica* (3). In 2016 researchers in Spain re-investigated the classification of the *C. zealandica* by Given and determined there to be several inconsistencies between the sample descriptions and therefore reclassified the grass grub beetle as *Costelytra giveni* (1). *C. giveni* is one of over 100 endemic beetles belonging to the Scarabaeidae family that can be found in New Zealand. Although *C. giveni* is a distinct species, it is believed that there might exist several subspecies of *C. giveni* (4), but further research is required.

C. giveni is holometabolic, meaning it has four major stages in its life cycle; egg, larva, pupa and adult beetle (5). The female beetles lay two to three batches of ~10-30 eggs each (5, 6). The beetles lay their eggs in the period of October – December approximately 20 cm below the surface. After ~3 weeks the eggs hatch and the first instar larva start feeding on roots and other larva, as grass grubs have cannibalistic tendencies (7). Around January – February the grass grub moults for the first time and enters the second instar stage while rising higher to the top soil to continue feeding. Around June-July, dependent on environmental factors such as temperature, the grubs reach the third instar. After several weeks, these third instar larvae start to burrow down to ~25 cm where they eventually pupate. The pupae will overwinter for up to 6 weeks before emerging as adult beetles and continue the cycle. A schematic of the grass grub life cycle is depicted in Figure 1-1.

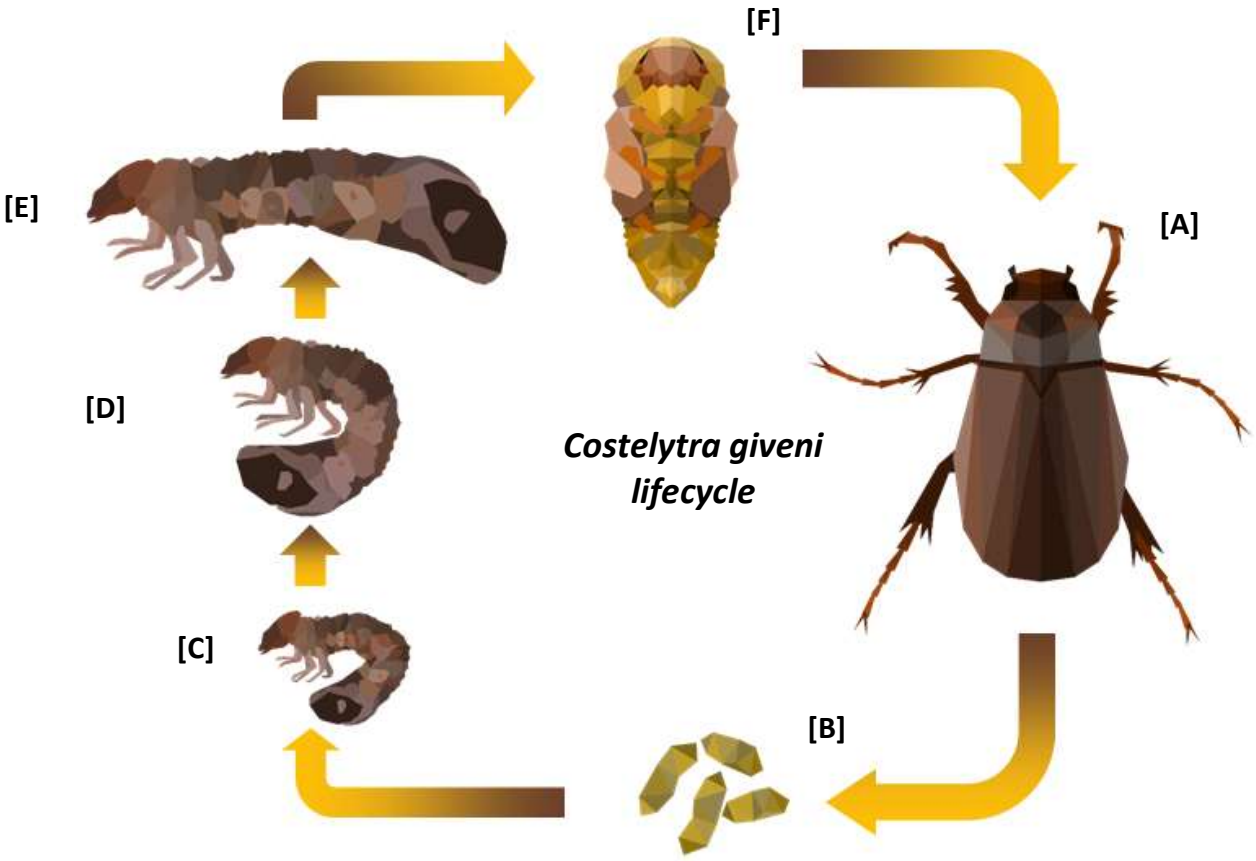


Figure 1-1; A generalised schematic of the lifecycle of the *Costelytra giveni.*A) Adult beetles are often seen in the period of October to December, B) a cluster of eggs laid round November – December, C) The first larva instar found around January – February, D) The second instar found around March – May, E) The third and last instar found around the month of June, F) The pupa can be found deeper in the soil between July and September. Variations can occur based on climate and location.

C. giveni was once mainly found in tussock in native bush and grasslands. Colonization of New Zealand by *Homo sapiens* however has caused major changes to New Zealand's flora and fauna, from the extinction of large land based birds such as the Moa by polynesian settlers (8) to massive loss of vegetation caused by both polynesian and early European settlers (9). Man made fire episodes alone are estimated to have resulted in loss of up to 50% of all of New Zealands native forests (10). The damage undertaken by humans to a fragile ecosystem resulted in a rapid adaptation response for many native species. When Europeans settled, substantial amounts of the remaining native flora was used for logging or removed to be converted into pastoral grasslands (11). The grass grub is one of the few species that easily converted to these new plant types and even prospered from their exoticism (12, 13). During the larval stage of the beetle's life, the grass grub stays below the soil and feeds on plant roots, most noticeably the nonindigenous pasture species such as *Lolium* spp. (rye grass) and *Trifolium repens* (white clover) (12, 14).

Grass grub populations often build up over a span of several years, becoming a major source of destruction to crops and pastures after four to five consecutive generations (15). It has become New Zealands biggest native pest to date (16). It was estimated in 2007 that grass grub had infested ~1M hectares of pasture soil (17). The destruction of plant roots, and in turn the viability of the plant and pasture as a whole, results in sizeable damages to the dairy and meat industry in New Zealand every year. In 1988 a paper was published by Grimont et al. (18) estimating these damages to be up to $100M NZD annually, but a recent study by Ferguson et al. has increased the estimated cost of damage to dairy and meat farms to $215 – $585M NZD annually (19). These estimates do not take into account the damages caused in agricultural crops that are also known to be affected by grass grub such as strawberry, rape, and turnips (20).

1.2 Chemical control of *Costelytra giveni*

Many synthetic pesticides have been used in the past or are currently used to control grass grub, some of which are shown in Table 1-1.

Table 1-1; An overview of some of the major synthetic pesticides used for grass grub control. [1] Perrot et al. (21), [2] East et al. (14), [3] Henzell et al.(22), [4]Laurent et.al(23), [5] Chapman et al. (16)

Organophosphates	Organochlorine	Pyrethroid	Noenicitinoids	Carbamate
Chlormephos[3]	DDT[1]	Deltamethrin[3]	Clothianidin[5]	Mexacarbate[1]
Chlorpyrifos[3]	Lindane[1]	Cypermethrin[3]	Imidacloprid[5]	Carbofuran[4]
Diazinon[3]				
Fensulfothion[2]				
Fenthion[1]				
Isazophos[3]				
Phorate[3]				
Terbufos[3]				

Often these pesticides have been ineffective (5, 22, 23) or are not equally effective in every New Zealand soil type (22). Apart from low effectiveness of some synthetic pesticides, a steady increase of pesticide resistance or tolerance has been observed in insects internationally (21, 23, 24). In grass grub, resistance to DDT has been documented (21) and population persistence has been observed under soil treatment with Fensulfothion, Isazophos and Lindane (23). There is also evidence that population resurgence can occur in pastures two years after treatment with Diazinon (25).

There are also indirect issues arising from the use of synthetic pesticides. These include: contaminating ground water (26) and rivers (27); lethal non-targeted effects on other organisms such as birds (28) and fish (29); harm to beneficial insects (30); reducing the fertility of the soil (31); negative health effects on farmers (32, 33) and consumers (34); and other effects (35). The main synthetic pesticide widely used in New Zealand for grass grub control is Diazinon. Despite the high effectiveness of Diazinon, this organic pesticide has been banned in many countries, including the US, due to its effect on other wild life such as birds (28) and even the human nervous system (36).

It has also been observed that Diazinon might have a natural grass grub bacterial disease suppressing property, and after several years, treated soil becomes more susceptible to infestation due to the removal of natural occurrences of pathogenic bacterial populations in the soil (25).

Due to increasing public concerns on the use of chemical pesticides, biological pesticides have slowly become a viable alternative as they are naturally occurring control agents as well as having the ability to co-evolve with their associated pest host. Continued research and development into these biological pesticides will reduce their production cost and eventually make them economically viable candidates to replace outdated chemical counterparts (37).

1.3 **Biological control of *Costelytra giveni***

Living microorganisms have been successfully used for centuries to help society in their day to day life, whether it is the use of yeast to make beer (38) or cheese (39), bacteria to make yoghurt (40) or to make filters based on reduced graphene oxide and bacterial nanocellulose for water purification (41), algae as a food source (42) and a catalyst to produce biofuels (43). It is therefore not surprising that microorganisms have become an important factor in the field of invertebrate pest control. Numerous microorganisms have been successfully commercialized and applied in the field to protect crops and fodder from invasive pests (37), although this process is slow and sometimes without much financial success due to poor performance or production difficulties (44).

Several biocontrol products are sold based on entomopathogenic nematode (EPN) species that have a mutualistic symbiosis with insect toxin encoding bacteria (45, 46). The most commonly sold and used species of nematodes are in the *Rhabditida*, *Steinernematidae* and *Heterorhabditidae* families (45). The process by which nematodes can affect invertebrate pests is by entering the host through any natural orifice. Once inside the host, the nematode will release its associated bacteria, most commonly members of the *Photorhabdus* or *Xenorhabdus* genera, that carry numerous antibacterial and anti-insecticidal pathogenic elements (47, 48).

The nematodes obtain their nutrients from the bacterial-caused decaying invertebrate host, then procreate and re-enter the soil in search of a new invertebrate host (49).

Fungal based entomopathogens comprise 19% of all biopesticide sales in 2010 (37). Some fungal entomopathogens come in the form of endophytes that inhabit a host plant and produce toxins that are harmful to invertebrates once ingested (50). Other fungal entomopathogens are free roaming and inhabit the soil or rhizosphere from where they can affect invertebrates directly (51). Apart from production of insecticidal toxin or stimulation of effectors produced by the plant host, numerous fungi also provide benefits to the plants they cohabitate with/in (52), such as provision of secondary metabolites that stimulate growth or protect against environmental factors, prevention of plant disease (53), regulation of defence responses (54) and many other qualities.

The largest portion of biopesticide products are bacterial based (37). There are numerous bacteria found with insecticidal properties, although only a small portion are commercially exploited as biocontrol agents (55). One of the most well-known entomopathogenic bacteria used as a commercial biopesticide is *Bacillus thuringiensis* (BT). The first BT strain found to have insecticidal properties was reported in 1901 by Ishiwata Shigetane, but the first published description and taxonomic classification of the bacterium was undertaken by Berliner in 1915 (56, 57). BT comprises a number of subspecies that produce unique parasporal crystals known as Cry toxins, with distinct Cry toxins having virulent effects on various invertebrate hosts (58-60). Genes encoding the Cry toxins produced by BT are often found on plasmids, some of which are lineage-specific (61) such as the Cry1A and Cry2A encoding pKur2, which is almost exclusively observed in Clade 2 *B. thuringiensis* isolates (61). Cry genes have also been shown to be transmissible between BT strains (62, 63), although not all *cry* genes are on Mobile Genetic Elements (MGEs) such as plasmids, and are therefore not transmissible this way. Through various methods, *cry* genes from BT strains have been incorporated into other plasmids, other BT chromosomes and even into plant cell lines, resulting in transgenic plants that then produce Cry toxins (64).

Although BT accounts for most the bacterial based biocontrol products sold, there are other bacteria that have been successfully commercialized as biopesticides. One such bacterial based biopesticide success story from New Zealand is that of the bacteria *Serratia entomophila* (65), which is the focus of this research.

1.4 *Serratia* spp.

The genus *Serratia* is part of the *Yersiniaceae* bacterial family (66). *Serratia* are non-sporeforming gram negative bacteria (18). These *Serratia* species are naturally occurring bacteria that are commonly associated with the grass grub (67). There are currently two *S. entomophila* biocontrol products being used in New Zealand to treat pastures against grass grubs (68) and one *Serratia proteamaculans* product is currently under development (69). A *S. entomophila*-based biopesticide product causes a chronic disease in grass grub called amber disease, also referred to in the past as "honey-coloured" disease (70), named for the amber-like discolouration in the larva gut that is associated with the disease (18). The symptoms include cessation of feeding and gut clearing and eventually, through starvation and associated weakening of the grass grub, sepsis and death of the larva. Three species of the *Serratia* genus are currently known to be able to cause this disease.

The first species of bacteria that now belong to the *Serratia* genus was described in 1931 by Grimes et al. (71) as *Aerobacter liquefaciens*. It took 40 years and several reclassifications, as summarized by Grimont et al. (72), before the *A. liquefaciens* was properly reclassified as *Serratia liquefaciens* by Bascomb et al. in 1971 (73). In 1978 Grimont et al. (74) found that the previously classified *Erwinia proteamaculans* had high biochemical and genomic similarities to *S. liquefaciens* and reclassified *E. proteamaculans* as *Serratia proteamaculans*. With the increased interest in the *Serratia* genus, and the advancements in classification methods Grimont et al. were able to do extensive work on classifying and reclassifying numerous members of the genus (74-76). In 1982 a new bacterial species was isolated in New Zealand and calssified as *Hafnia alvei* by Trought et al. (70). The *H. alvei* was described as causing a "honey-coloured" disease in grass grubs. In 1988 this *H. alvei* was officially classified as *Serratia entomophila* by Grimont et al. (18). One of the isolates analyzed in the Grimont et al. paper (18) was an isolation from 1981, obtained by T. Jackson, from the Fairton area in Canterbury, New Zealand, designated as A1. The *S. entomophila* A1 strain was one of the strains causing the previously named "honey-coloured" disease, that has since been renamed as amber disease. This *S. entomophila* A1 strain, later dubbed A1MO2 (77), was known to have pathogenic properties towards grass grub. In 1986 the A1MO2 strain was put through the process of commercialization in collaboration with the Ministry of Agriculture and Forestry (MAF) and Monsanto NZ Ltd. (65). Due to several problems and disappointing sales, Monsanto withdrew from the collaboration (17). Eventually, in 1990, the A1MO2 strain was sold under the name Invade™ by Coated Seed NZ Ltd. (65, 78, 79). Invade™ is a liquid formulation that is injected into the soil using a modified seed drill system. The liquid formulation was produced by Industrial Research Ltd. in 500 L batches and quality control was performed by AgResearch Ltd. (78). The liquid product could be kept in a 4°C storage for about 20 weeks before substantial loss of viability was observed (78).

In 1991 Jackson et al. published a review where they analyzed the pathogenicity of several *Serratia* isolates that they had collected by that time (80). In this review they found 46 out of 82 analyzed *S. entomophila* isolates to be pathogenic, approximately 56%. In the analyzed *S. proteamaculans* isolates they only found 18 out of 45 to be pathogenic, this translates to 40%. None of the other species or isolates from other genera were found to be pathogenic. Although the number of analyzed isolates was small, it was assumed that this number could be translated to the entire *Serratia* population of New Zealand.

Further research into the A1MO2 strain undertaken by Glare et al. (81) found the mechanism behind the pathogenicity was correlated to a megaplasmid (i.e. plasmids >100 Kb in size (82)), estimated at the time to be 105 Kb in size. In 2011 however Hurst et al. sequenced the original pADAP plasmid and found it to be 153,404 bp in size (83). Claus et al. (84) and Grkovic et al. (85) further confirmed the relationship to pathogenicity and determined the genomic regions and plasmid profiles of different isolates by restriction digestion of major areas of the plasmid. It was not until 1996 that Glare et al. (86) showed that the megaplasmid, now designated as pADAP (for amber disease associated plasmid), was the only element needed to induce amber disease in grass grub. They proved this by performing bioassays with *Serratia* species and even several non-*Serratia* strains including *Klebsiella* sp., *Enterobacter agglomerans* and *Escherichia coli* that were transformed to now contain the pADAP megaplasmid. All bioassays with these plasmid containing strains caused amber disease, confirming that the host bacterium is not important to induce pathogenicty.

Between the time A1MO2 was on the track of commercialization from 1986 and being used in the field in the 90's, the production of Invade™ would occasionally suffer from major production problems. One major problem initially overlooked in the early stages when the presence of the plasmid had not yet been discovered, was sudden loss of the pathogenic plasmid during production (78). Once the plasmid importance was uncovered, this problem could be adressed, however it did mean that pastures where the Invade™ product had been applied, needed to be retested for presence of A1MO2 with plasmid at one point. Other problems included sedimentation of bacteria when in storage, resulting in some fields not having gotten proper coverage of bacteria, the type of drill needed for injection of the liquid was not usable in all terrains, and several other problems were observed with the liquid Invade™ product (78). Also later analysis of the shelf life of Invade™ brought down the estimated 20 weeks to only 7 days of effective storage, substantially reducing the cost effectiveness of mass production of the product (87).

Around the turn of the 21st century a newer formulation of the A1MO2 product was made and is currently sold by BioStart™ under the name Bioshield™ (87, 88). Bioshield™ is typically delivered as an inert zeolite granule, deposited onto the ground and absorbed into the soil, assisted by irrigation. Injection into the soil by a regular seed drill has been mentioned as a more optimal approach to obtain proper coverage. This new formulation is more stable and could potentially be kept for an estimated 180 days at ~20°C, reducing the cost for cooled storage, and eliminating the 'produce on demand' needed for Invade™ (due to its 7 day shelf life) (87). The A1MO2 strain was eventually replaced with the phage resistant 626 strain in the production of Bioshield™ to enhance product viability in the field.

The larvae affected by amber disease typically die within in 1 to 4 months (89). Bacterial breakdown of the larval cadaver, persistence of bacteria in the soil (70) and the cannibalistic nature of the grass grub at high population density (7), ensures that the bacterial population is released back into the soil to form stable protection for the next season (17). *Serratia* populations can steadily build up over several years in a pasture as grass grub numbers rapidly increase. This increase of both the target insect and the biocontrol agent leads to a grass grub population collapse around five years after *Serratia* applications (68). It is unknown if this collapse can be solely attributed to the disease background reaching a critical point or whether other effects, such as depletion of resources for the grass grub population to sustain itself, play a significant role as well.

Where during the 1990's Invade™ could cause a 40-50% population decline in the season of application (65), more recent studies show that its effect might be waning or have been overestimated, with one study showing 76% of grass grubs surviving in treated pastures (5) and another showing an increase in grass grub population after the second year of treatment in certain pastures (25). The chronic nature of *S. entomophila* induced amber disease coupled with the variable effectiveness and the five year build up needed for successful population collapse versus the quick and rapid death assured by competing synthetic pesticides has stimulated research into finding alternative *Serratia* strains with better and faster grass grub control characteristics. *Serratia* also contains a species called *S. proteamaculans* with strains that have naturally occuring grass grub insecticidal activity encoded on pADAP-like plasmid (90). Some of these *S. proteamaculans* strains can also induce amber disease in *Pyronota* beetles, namely *P. festiva* and *P. setosa* (69), termed kēkerewai by the Māori and colloquially referred to as Manuka beetle as they co-habitate areas filled with manuka trees (91, 92). Manuka beetle is not yet a major national pest, but has been shown to be able to become a pest locally in specific regions, most notably in the West-Coast in soil that is converted to dairy pasture through a process called "flipping" (93).

Although *S. entomophila* and *S. proteamaculans* both belong to the same genus, DNA-DNA hybridization studies have shown that *S. proteamaculans* appear to only have 32% to 40% of genomic DNA sequence similarity to *S. entomophila* (18) although recent chromosomal sequence comparisons undertaken by Vaughan [unpublished data] show *S. proteamaculans* has a 84% sequence identity to *S. entomophila* chromosomes. Apart from certain *S. proteamaculans* isolates having a wider host range, some also have a quicker pathogenicity response, killing within 12 days as opposed to 1-3 months. One example of a *S. proteamaculans* with a "hyper-pathogenicity" pathotype is AGR96X (69). An example of both the chronic and the 'hyper-pathogenic' disease phenotypes are shown in Figure 1-2. In addition to the *S. entomophila* and *S. proteamaculans* isolates, this study identified two isolates of *Serratia liquefaciens* that are able to induce the chronic amber disease in grass grubs, which had not been previously reported.

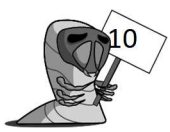
Figure 1-2; Overview of the two known disease states caused by *Serratia* spp. on *C. giveni* larvae. In the top-middle is a picture of a healthy grub, the hindgut is dark-brown, and the body has a grey-like hue due to the soil and food present in its semi-translucent body. On the bottom-left is a grub that died from a "hyper-pathogenic" *Serratia* strain that caused rapid death after several days. On the bottom-right is a grass grub suffering from amber disease; this grub will no longer feed, regularly void its gut, and eventually die months later due to malnutrition.

1.5 The nature of plasmids

pADAP, the host of the main genetic components behind the pathogenicity of these *Serratia* isolates, is a plasmid that was first found in a strain designated as A1MO2 (81). Bioassays showed that plasmid free isolates of the bacterium were avirulent (85). Previously analyses undertaken on the pathogenicity of isolates, by Jackson et al. (80) and later again by Dodd et al. (94), showed that approximately half of the assessed isolates were pathogenic. These results led to the general assumption that half of the field isolates carry pADAP.

Plasmids are circular DNA molecules, easily shared from cell-to-cell. This is done in two steps, a “sex” pilus is formed that attaches to a recipient cell and contracts (95), such as the pADAP encoded Type IVa pilus (96). After contraction of the pilus, a conjugative pore can form between the two cells that allows for single stranded DNA transfer from donor to recipient (97). But plasmids can also be passed down from parent to progeny. Plasmids contain an origin of replication referred to as the ori, that often have an ORF that encodes for some type of replication protein (Rep) needed for replication of the plasmid, although not all plasmids have one (98). Often plasmids are also classified in groups called incompatibility groups (Inc), based on their ori (99, 100). This method is based on the idea that, in some cases, two replicons cannot be maintained simultaneously, and only one plasmid will be able to propagate.

Plasmids are often replicated through one of two ways: theta-type replication and rolling-circle replication (98, 101). Theta-type replication is a method of plasmid replication where two single stranded initiation sites of the plasmid form hairpins through interaction with a replication gene. The hairpins pull both strands apart, forming a loop. From this loop, bidirectional extension of the single stranded DNA (ssDNA) is performed, resulting in two copies of double stranded DNA (dsDNA each containing a single strand of the original plasmid and one complemented strand. Rolling-circle replication, where one strand of the DNA is nicked and replication happens through strand displacement, separating lagging strand of ssDNA while the leading strand of ssDNA is regenerated into dsDNA, afterwards the ssDNA plasmid copy is also converted back into dsDNA, resulting into two plasmid copies (101, 102). Rolling-circle replication can also occur during plasmid transfer from one cell to another, also known as conjugation. It does this by pulling the lagging strand into a conjugative pore or a secretion system and exporting it out of the donor cell, converting the lagging strand into dsDNA happens in the recipient host if there is one (103).



Conjugation of plasmids is one of the easiest modes of transferring novel genetic material, also referred to as Horizontal Gene Transfer (HGT), into other cells. HGT is an important evolutionary trait, especially for prokaryotes, to allow for rapid evolution through acquisition of novel genetic material by means other than natural reproduction (104).

Bacteria can acquire novel genetic material through of several mechanisms; Transduction per mediation of viral or phage like particles containing DNA packages (105); Transport via mediation of membrane vesicles (MV) (106); Acquisition through mediation of gene transfer agents (GTA) (107); Transformation of foreign DNA or RNA directly into the host cell (108); Bacterial conjugation of DNA molecules and plasmids (109).

Conjugative plasmids have been shown to be a key driver in the rapid adaptability of bacteria to environmental factors. Factors such as resistance to high concentrations of heavy metals (110), circumvention of host immune responses to ensure survival (111), resistance to antibiotic elements (112, 113), resistance to pesticides (114), supplementing the cell with additional mechanisms to convert or construct novel metabolites (115), decrease microbial competition through secretion of anti-microbial compounds (116), and many other features can be encoded on plasmids. Plasmids have therefore been considered a core element of a population's "supergenome" by some researchers, as they are part of an exchangeable set of genes, shared by large populations of different bacterial strains and species (109, 117).

Plasmids come in all sizes, some even larger than the size a functioning prokaryotic chromosome. The smallest primary bacterial chromosome has been documented to be that of *Carsonella ruddii* and is only ~160 kilobases in size (118) whereas one of the largest documented plasmids is the 1.8 Mb linear plasmid pSCL4, found in a *Streptomyces clavuligerus* strain (115). Plasmids are often regarded as accessory genomes due to the fact they are easily transmissible, rarely encode housekeeping genes and can often be discarded without consequences to the functioning of the host cell. There are, however, some rare cases of core gene-containing plasmids, referred to as chromids (119), or secondary chromosomes (120), and they are the prime example that blurs the line between a chromosomal host and its genetic parasites.

1.6 The burden of plasmids

Plasmids often are equipped with a slew of genes that regulate the expression and replication of the plasmid itself and the genes encoded on it, and can therefore impose a metabolic burden to the cell (121). Not only the replication and conjugation of the plasmid are considered a metabolic burden but also nucleotide bias, the depletion of amino acids (AA) during translation of genes, cytotoxic effects of novel proteins, triggering of cell SOS-signal pathways, the acquisition of transcriptional and translational resources from the host, and many other costly characteristics of plasmid carriage can affect the fitness of a cell (121-123).

This metabolic burden is why plasmids are often considered to be parasitic in nature and are sometimes referred to as selfish genetic elements (SGEs) (124, 125). Although it is hard to put a number on the actual metabolic drain a plasmid can have on its host, there are studies reporting the differential expression between hosts with and without a plasmid. Shintani et al. (126) described only ~1 – 4% of chromosomal genes are differentially expressed between pCAR1 plasmid bearing cells vs non-bearing cells, mostly upregulated. However, Harrison et al. (127) found that mutation in the *gacA*/*gacS* genes, a two-component regulatory system responsible for regulating a large amount of secondary metabolites and antimicrobials (128), on the pQBR103 plasmid in *Pseudomonas fluorescens* resulted in downregulation of ~17% of chromosomal and plasmid genes. This means that without that mutation the plasmid would be a transcriptional burden through upregulating ~17% of chromosomal genes, in addition to genes it is already upregulating by being present in the host (127).

Millan et al. (129) performed a more comprehensive study on the *Pseudomonas aeruginosa* PAO1 strain, and the fitness effects that plasmid carriage has on it. In their study the authors showed that plasmid genes are expressed at a higher rate than those of the chromosome, most notably antibiotic resistance genes, with the exception being genes related to the conjugative machinery on the pAKD1 plasmid, which were repressed. In total six PAO1 strains, carrying different plasmids, were tested over a range of different carbon sources and it was shown that only two plasmids had an average growth benefit, whereas the other four had mild to severe negative effects on the growth rate of the PAO1 strain compared to plasmid free cultures. Strains carrying the plasmids also showed a substantial differential expression on metabolism and secretion genes in the chromosome, essentially showing that the plasmids severely altered their host's metabolism.

A study undertaken by Hall et al. (130), assessing several mercury resistance pQBR family plasmids, showed a clear fitness benefit for strains void of any plasmids in the absence of mercury. What they also showed is by adding increasing amounts of mercury to medium, there is a tipping point where the mercury resistant plasmid bearing strains start having a substantial growth/survival benefit. This is a great example that these genetic elements can be considered parasitic in certain conditions but do allow the host to transition into new environmental niches.

As demonstrated by Hall et al. (130), some plasmids also seem to fulfil symbiotic functions in that they provide their hosts with novel mechanisms to survive xenobiotic contaminants or ecological factors but they also allow for acquisition of novel genetic material without disturbing the evolutionary pressure under which the chromosome operates, allowing for plasticity of the host (131, 132).

Plasmids can act as a genetic capturing mechanism that allows for more lenient selective pressure than is expected from the chromosome, for example, plasmids can contain genes with a lower expression rate, thus having less evolutionary pressure for translation efficiency for these genes (133-136). Plasmids also often contain less or even no essential genes that are dispensable and thus can tolerate higher rates of mutation (136, 137). These genes will also suffer lower rates of evolutionary pressure if they interact less with other proteins and have fewer functions (136). Plasmids can essentially operate as recombination hotspots (RHS) for the creation of novel genes and protein products by allowing vastly higher rates of acquisition, insertion, deletion, duplication, reshuffling and recombination of their genetic make-up than the chromosome could (138). In addition to these attributes, through acquiring novel pathogenic systems, metabolic enzymes or other characteristics, plasmids can facilitate the transition of their hosts into new ecosystems (139). A great example of this is the acquisition of the pADAP plasmid by *Serratia* species, that allowed them to invade, debilitate and exploit the grass grub as a novel food source.

1.7 **Plasmid persistence**

Due to the parasitic characteristics of these mobile genetic elements (MGE), it is often assumed that the host would lose the plasmid over time due to their sometimes-unfavourable selection pressure and the resulting decrease in its host's fitness. However, plasmids are found throughout the prokaryotic kingdom, often highly abundant within populations. From this fact it is apparent that plasmid loss is a rare event as exemplified in a study undertaken by Glare et al. (81), who had to incubate pADAP bearing *Serratia* isolates under growth inhibitory temperatures of 40-42°C for 7 days, resulting in 99% cell mortality in the process, in order to get any plasmid cured colonies.

The probability for a cell to become plasmid free from a high copy plasmid is very low. Sorensen et al. (132) estimated that for the cloning vector pBR322, which has ~21 copies per cell, the complete loss of all plasmids only happens in 1 out of every 1M cells (132). Based on statistical models, some researchers estimate that it might not even be possible for a bacterial population to lose plasmids at all (140) and that emergence of parasitism of genetic elements such as plasmids and their persistence is actually an expected driving force of evolution (141).

Some plasmids employ a toxin-antitoxin (TA) mechanisms to maintain their presence in a cell (142, 143). Several different types of TA mechanisms exist, some causing post-segregational killing (PSK) (144), with others working more as antimicrobials on other cells that do not contain the plasmid (145, 146), and some TAs can even cause persister states where the affected cells become metabolically inactive if they do not maintain the plasmid with the anti-toxin (147, 148).

Other systems build up toxins within the cell ensuring either vertical transmission of the plasmid or PSK (143). Variations of these systems exist, and TA's are very abundant in prokaryotes (143), but not all plasmids contain TA systems, and not all TAs are present on plasmids.

The duality of plasmids being a burden that should be negatively selected for directly contradicts the findings from many studies in this field that determined that plasmids are abundant, widely distributed and highly diverse (149-153) and it appears that plasmid loss does not happen often or easily (132) (140) or only happens in a very limited rate under very harsh conditions such as extreme stress (81). The duality is referred to as the 'plasmid paradox' (154).

Despite having parasitic characteristics, plasmids have been shown to be a key factor in the rapid adaptability of bacteria to new environmental factors such as high concentrations of heavy metals (110), circumventing host immune responses to ensure survival (111), resistance to antibiotic elements (112, 113), resistance to pesticides (114), supplementing the cell with additional mechanisms to convert or form novel metabolites (115), decrease microbial competition through secretion of anti-microbial compounds (116) and many other features.

They act as a capturing mechanism that functions under very low evolutionary pressure compared to the primary chromosome (131), making it easy to acquire novel genetic material without disrupting functionality of its bacterial host. Jain et al. (155) estimated that HGT, not just through plasmid transfer, has accelerated genome change of prokaryotes by a factor of at least 10^4, if not higher.

There are also studies that show that co-evolution of plasmids and the primary chromosome can result in lower metabolic cost, thus negating the parasitic traits of the plasmid almost completely (121, 127, 156-158). One method of co-evolution is by suppression of plasmid replication by the host (127). There are several ways by which this can be undertaken such as plasmid dimerization, called "handcuffing". The subject of plasmid regulation through handcuffing in the context of the pADAP plasmid, will be further investigated in Chapter 4.

Common means of replication suppression is through transcription of antisense small noncoding RNAs (snRNAs) that can disrupt several processes related to replication (159). These include: binding to mRNA formed by the *repE* gene in the pAMß1plasmid in *Enterococcus faecalis*, thus causing premature termination of transcription of the replication gene; the binding of antisense RNA to the Shine-Dalgarno (SD) Sequence of the *repE* gene in the plasmid pMV158 in *Streptococcus pneumoniae* gene, thus blocking ribosome binding and further translation; the binding of antisense RNA to the leader peptide containing residues for RepY and RepZ associated with the pColbP9 plasmid in *Escherichia coli*, preventing it from forming a pseudoknot, a secondary structure that is needed to expose the SD sequence; several other antisense mechanisms are known (159). These kinds of down regulation of plasmids can alleviate the metabolic burden on the cell.

1.8 pADAP and its features

The pADAP plasmid is a self-replicating, conjugative genetic element, associated with three soil bacteria of the *Serratia* genus in the family of Yersiniaceae (66), *S. entomophila*, *S. proteamaculans* and S. *liquefaciens*. This study uncovered that some non-endemic isolates, believed to carry a pADAP plasmid, turned out to not carry pADAP but instead carry a different megaplasmid of similar size. The pADAP plasmid has therefore, as yet, not been isolated from anywhere other than these three *Serratia* species, or outside of New Zealand, although homologous elements of the pADAP plasmid have been found on other plasmids (90).

The pADAP plasmid consists of a core backbone containing a region of replication-associated genes and a region of genes correlated to transcription and regulation of a conjugative machinery (Figure 1-5), making it a self-replicating, self-conjugating plasmid (83). Besides the backbone, the first pADAP sequence was found to contain a fimbria encoding region, designated as *Serratia entomophila* fimbriae (Sef) region, not to be confused with the *Salmonella enterica* fimbriae (Sef) cluster published three years prior (160). The mechanisms behind the pathogenicity of pADAP are encoded on the two putative pathogenicity associated islands (PAIs) (83). The first PAI encodes for a bacteriophage-like particle called Afp, short for Anti-feeding Prophage (161), and the second PAI encodes a pore-forming system called Sep, short for *Serratia entomophila* Pathogenicity (90). PAIs are genetic regions that encode pathogenic determinants and the PAIs on pADAP are flanked by genes encoding hypothetical proteins and transposon-like elements, which suggests that these regions at some point have been able to transpose themselves in, and potentially out of, the plasmid or genome of origin, indicative of HGT (162, 163). A genetic map of the pADAP from the Genbank accession [NC_002523] (83) can be seen in Figure 1-3.

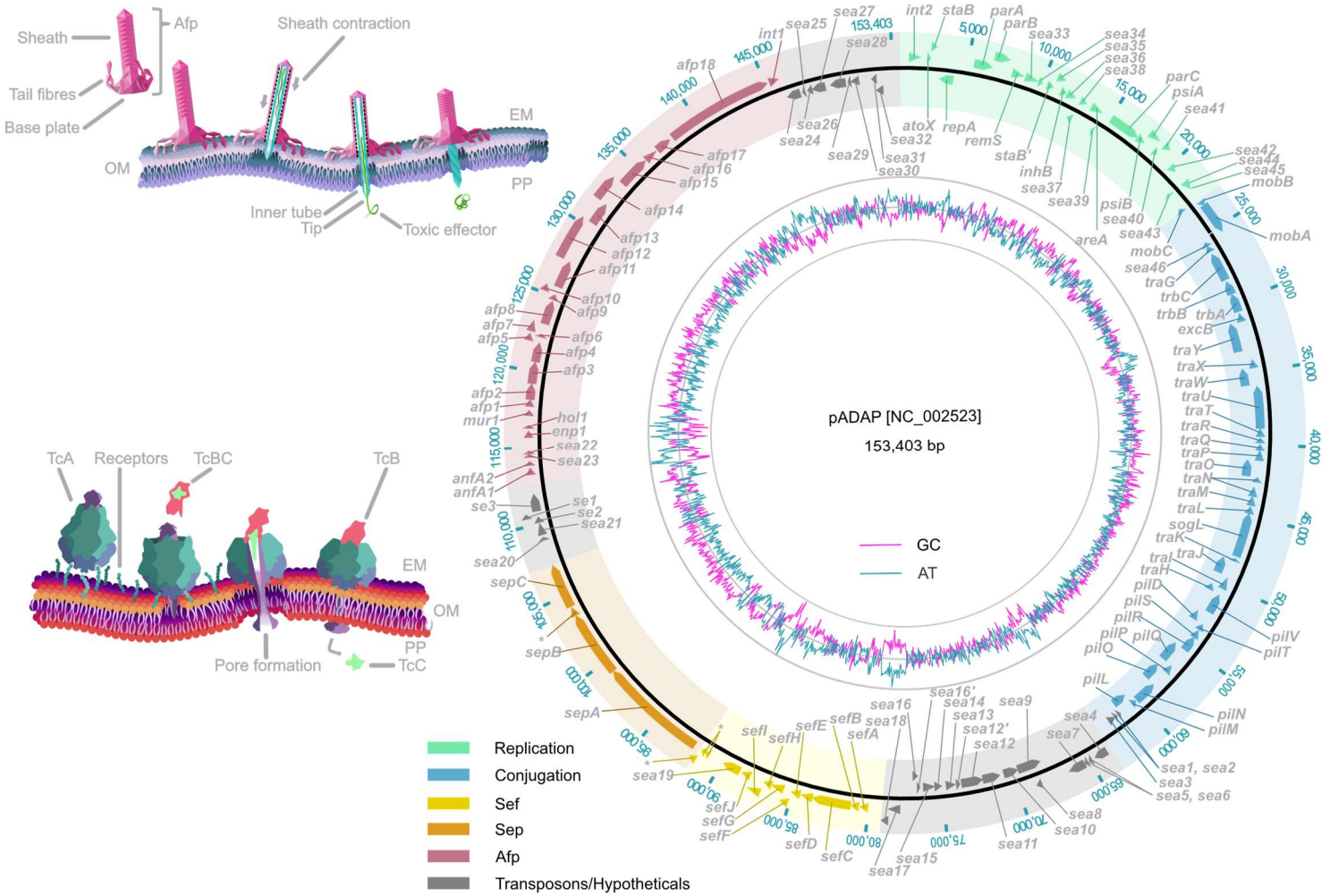


Figure 1-3; Schematic of the pADAP reference plasmid from *Serratia entomophila* strain A1MO2. This figure shows the features of the pADAP plasmid as described by Hurst et al. (83), as well as a GC plot and an illustration of Afp (top left) and Sep (bottom left). Green denotes the replication region, the blue features are genes associated with the conjugation, Sef fimbrial region is marked in yellow, in brown the Sep PAI and the Afp PAI features are highlighted in burgundy. Grey regions are clusters of hypothetical and transposon like elements. The sequence was obtained from NCBI under accession number [NC_002523].

MGEs such as pADAP often seem to fulfil a symbiotic function in that they provide their host with novel mechanisms to survive xenobiotic contaminants or ecological factors, such as toxic environments, or in the case of pADAP, encoding mechanisms for facilitating the colonization of the gut of the larval stage of the *Costelytra giveni*. A schematic of the lifecycle as well as the pathogenic elements causal to amber disease can be seen in Figure 1-4.

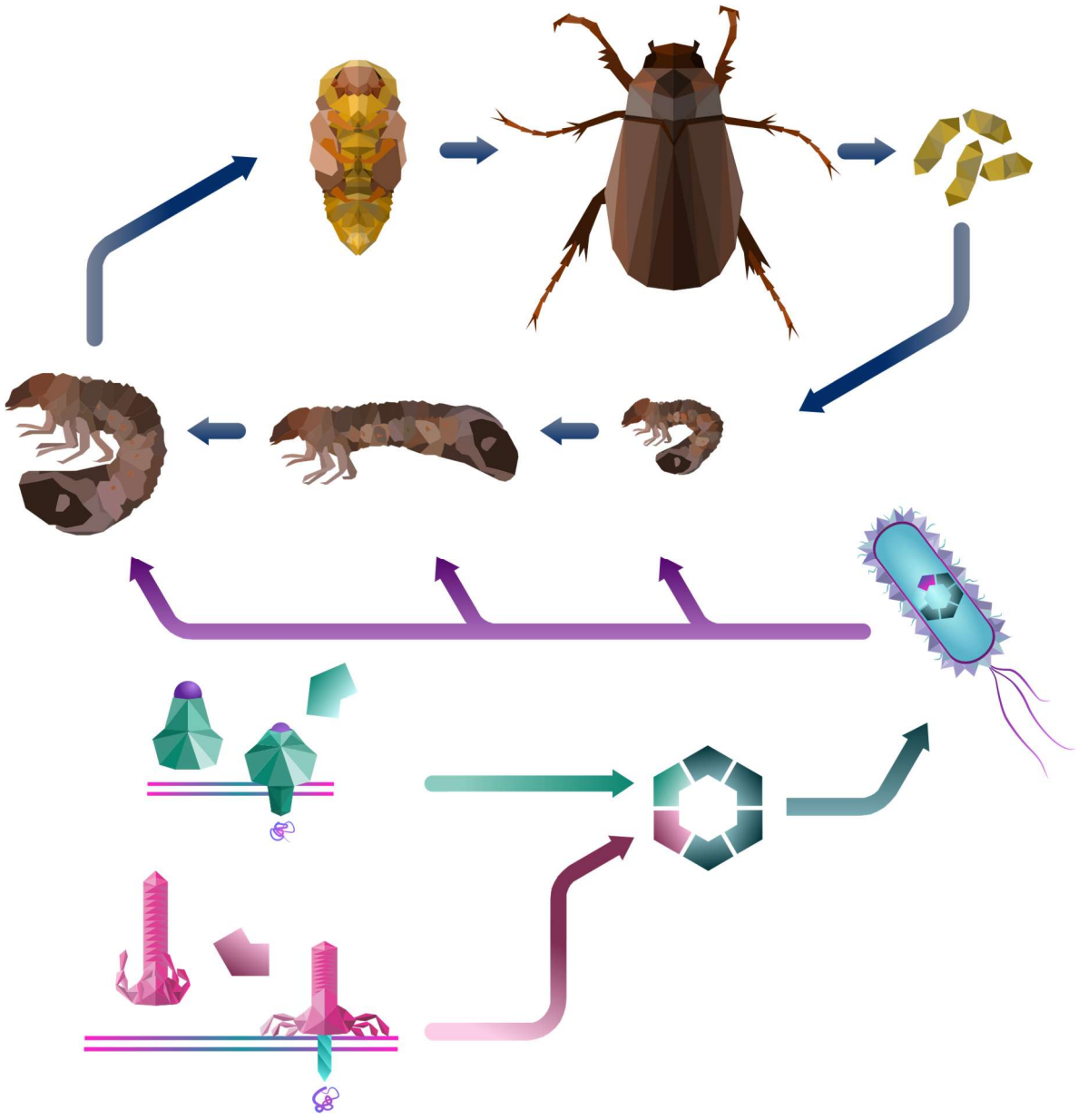


Figure 1-4; A figure showing the lifecycle of the *Costelytra giveni* beetle, and the bacterial toxins Sep and Afp, that cause amber disease. The grass grub has four major life stages, adult beetle, egg, three instars of larvae and pupae stage. *Serratia* bacteria containing the pADAP plasmid can affect the larvae stages of the *C. giveni*. It can do so by producing two toxic effectors, SEP and AFP. Sep is a pore-forming secretion system that is part of the ABC Toxin complex family and causes gut clearance and reduction of proteases. Afp is a tailocin secretion system that can adhere to cells of the grass grub and inject toxic effectors into the cell and causes anti-feeding behaviour though a yet undetermined mechanism.

1.9 **Conjugation**

The pADAP plasmid is considered a conjugative plasmid because it encodes for an extrachromosomal conjugative machinery. The conjugative region is around 40 Kb in size and starts from *mobC* and approximately ends at *sea1*. This region starts with three genes belonging to the *mob* family (164). The *mob* genes, short for mobilization genes, are required to prepare the plasmid for transfer. MobA contains a relaxase domain known to be associated with DNA nicking, and a primase domain that is required for DNA replication (165). MobB increases the proportion of plasmid molecules nicked by interacting with transfer genes (166). MobC has high similarities with proteins involved with the interactions between relaxosomes and the secretion system through which the plasmid will be transferred (164).

Apart from the *mob* gene family, the conjugation region comprises 17 genes belonging to the *tra* family (167). Tra is short for transfer, the translated products of these genes are important for nicking and linearizing plasmid DNA at the origin of transfer through a complex of DNA-binding proteins called the relaxosome. Part of the relaxosome consists of mating pore formation (Mpf) proteins (168) that construct a transmembrane pore called the mating pore (97, 169, 170). This pore allows macromolecules and DNA to pass the inner membrane (171, 172). This Mpf is shown to be coupled to a Type IV secretion system (173).

A Type IV pili (T4P) (174), belonging to the Pil, short for pilin, family, is transcribed on 11 open reading frames (ORFs) on pADAP. This Pil system is grouped with other Type IV pili such as the F pilus also called the F factor or sex pili (175) and is a type IV secretion system (95). These T4P complexes allow for attachment to a recipient cell, upon which the pili contracts allowing for cell-to-cell contact (95, 176). But the T4P also enables movement of the cell through twitching motility, enables microcolony formation by adhering to remote cells, and allow cell transformation by internalizing extracellular DNA (177-179). Transfer of DNA from donor to recipient across the F pilus has also been shown to happen between cells up to 12 µm apart (180).

Transport of the linearized plasmid to another cell as a means of asexual transfer of genetic material is referred to as HGT. Models suggest the Mpf transports the DNA cross membrane upon which either direct cell-to-cell transport of DNA can take place (95, 172), or transport across the pili happens (180), one directional from donor to host.

As there is evidence supporting both modes, the specific preferred mode remains unresolved (168, 181). A schematic of the region as well as an indication where it is situated on the plasmid can be seen in Figure 1-5.

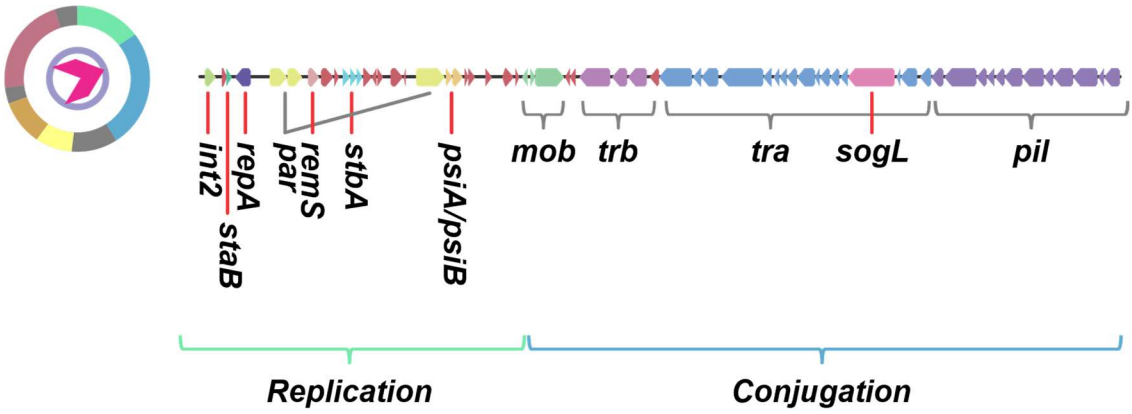


Figure 1-5: pADAP backbone, consisting of a replication and a conjugation region. Relative position of this region on the plasmid is indicated by the centre pink arrow on the representative plasmid map in the top left corner. The conjugative region consists of three *mob*, mobilization genes, three *trb genes*, seventeen *tra* transfer genes and thirteen pil genes.

By having this conjugative region on pADAP, the plasmid can readily be transferred to other cells without the need of chromosomally encoded genes, thus potentially increasing the number of cells carrying the Sep and Afp virulence determinants and increasing biological control of grass grub in a natural setting. The pADAP conjugative region is similar to the conjugation region on the *Salmonella enterica* serovar Infantis pESI and pR64 (182)

1.10 *Serratia entomophila* fimbriae

The pADAP plasmid encodes for a Type 1 fimbria (83, 183). The region was designated as the Sef operon for *Serratia entomophila* fimbria. The Sef encoding region contains ten Sef genes, *sefA-sefJ*, as well as the *sea19* ORF (Figure 1-6).

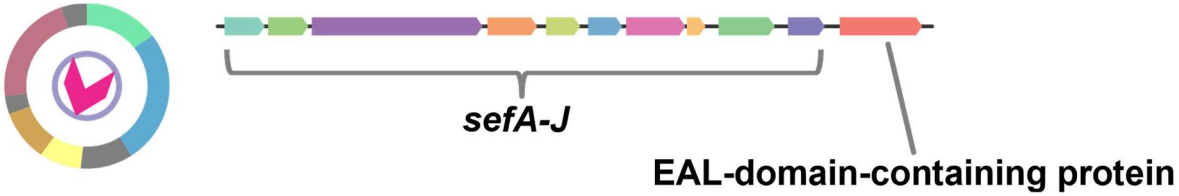


Figure 1-6: A schematic of the Sef region of the pADAP plasmid. The Sef region consists of 10 genes transcribing a fimbrial unit, and one hypothetical gene that contains an EAL domain. EAL domains are associated with biofilm regulation and does not appear to be needed for fimbriae formation.

Type 1 fimbriae are mostly associated with the ability of *E. coli* cells ability to adhere to mammalian cells in the urinary track such as that of humans (183, 184), mice (185) and swine (186). These Type 1 fimbriae have also been correlated to evading extracellular antibiotic treatments, and immune responses of the host by helping bacteria with adherence to niche locations inside the host body (187). The Type 1 fimbriae have also been shown to be an important factor in colonization and proliferation of *Xenorhabdus nematophila* in organisms such as nematodes (188) as well as contributing to pathogenicity towards insects such as cotton bollworm (*Helicoverpa armigera*) (188). The Sef region is directly followed by the Sep region.

1.11 *Serratia entomophila* pathogenicity toxin complex

The Sep particle is a member of the ABC toxin complex (Tc) family (189). Toxin complexes belonging to this ABC Tc family consist of three components, simply termed A, B and C-like components (190), that form a syringe like structure capable of injecting protein based toxins into the recipient cell (47). The Sep encoding region is displayed in Figure 1-7

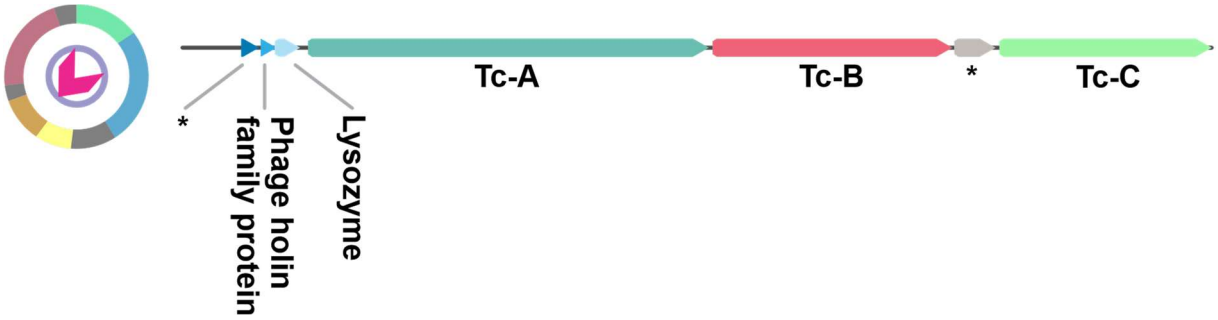


Figure 1-7: A schematic of the Sep region of pADAP. 5' of the Sep encoding ORFs are a holin and lysozyme, needed to release the particles by lysing the cell.

The Sep operon is highly similar to the TCs found in the nematode associated bacterium *Photorhabdus luminescens* (190, 191) and suggests acquisition through HGT at some stage (192). A direct homolog of the Sep operon has also been found in *Yersinia frederiksenii* strain 49 (90), which, like *Serratia*, is also part of the Yersiniaceae family (66), confirming that these PAIs at some point were transmitted horizontally. This homolog will be further discussed in Section 4.10.

The three Sep components, SepA, SepB and SepC together form the Sep complex (190) and are required for the complex to form and function properly (193, 194). In these ABC-complexes, the Tc-C component is the toxic effector (195) which in the case of *sepC* is confirmed by its lower ratio of sequence conservation between pADAP plasmids compared to *sepA* and *sepB*, as well as lower GC%.

A lower GC% and a higher mutation rate are features often shared by pathogenic regions, also referred to as Recombination Hot Spots (RHS) (196) .

After translation of the ABC components, the C terminus of SepC is enveloped by SepB (197). The SepB component has a dual function acting as both a connector between SepA and SepC and as a potential protective barrier that protects the host bacteria from the SepC –C terminus cytotoxic region. Purified Sep causes gut clearing in grass grub (190, 198). A schematic of the Sep complex and its mechanistic action is depicted in Figure 1-8.

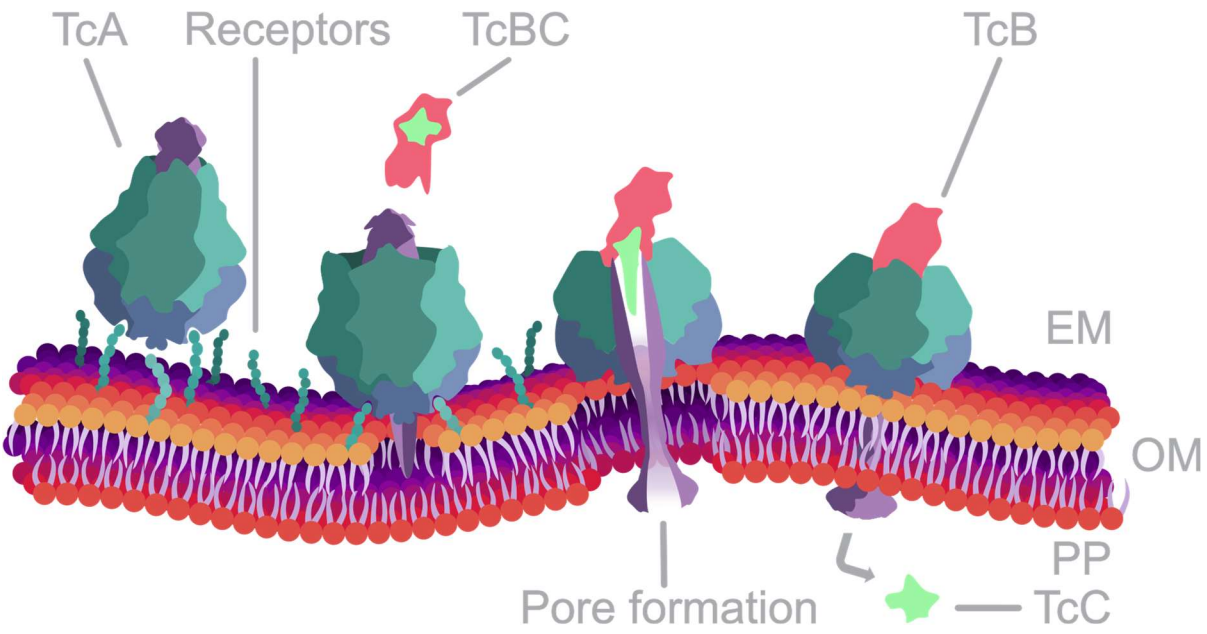


Figure 1-8; A simplified Illustration of the *Photorhabdus luminescens* ABC Toxin complex. This figure shows the mechanism with which the Tc affects target cells. This Tc complex is considered similar to the mechanism of Sep/Spp. The TcA unit forms a pore-forming unit, that upon interaction with receptors on the OM of a target cell, undergoes conformational change. An inner tube domain gets pushed through the OM to form a pore. One the pore is established, a TcBC complex can bind to the now changed TcA complex. The TcC is then transported into the periplasm.
Illustration is based on a model by Gatsogiannis et al. (48).

The SepB component connects the SepA and SepC component and acts like a shell that envelops the SepC toxin, preventing any interaction between the host cell and C-terminus cytotoxin region of SepC (197), similar to the mechanism in *P. luminescens* (48). Based on the known function of Sep orthologue found in *P. luminescens*, as described by Gatsogiannis et al. (48), the SepABC complex, once formed, is able to force a pore through a target cell membrane and translocate the toxic effector into the target cell as depicted in Figure 1-8.

The closely homologous toxin complex found in *S. proteamaculans* pADAP plasmid pU143 was dubbed Spp for *Serratia proteamaculans* pathogenicity (83). This system also consists of three genes, *sppA*, *sppB* and *sppC* and most likely forms the same complex as Sep and the *P. luminescens* ABC Toxin complex.

Post challenge, both the Sep and Spp cause voiding of the gut in the *C. giveni* larvae (161, 190) which severely reduces the concentration of major *C. giveni* gut digestive enzymes such as trypsin and chymotrypsin (199, 200).

1.12 **Anti-feeding prophage**

The Afp particle is transcribed as 18 ORFs Figure 1-9. The translated products of *afp1 – 16* forms the observed Afp particle (161). The Afp1-16 proteins share high AA similarity to six *Photorhabdus* virulence cassettes (PVC), present in some pathogenic species of nematode associated *Photorhabdus* bacteria (191).

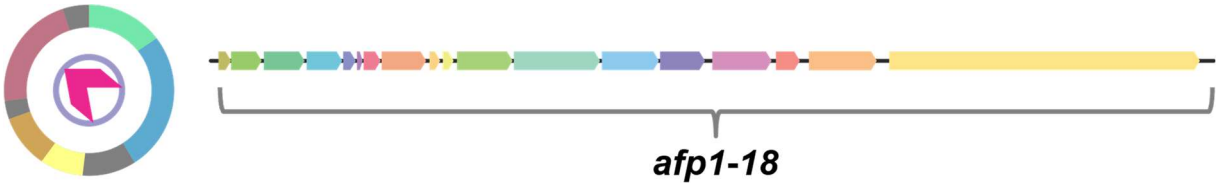


Figure 1-9: A schematic of the Afp region of pADAP. The region contains 18 ORF that make up the Afp apparatus. *afp1* and *afp5-7* are genes involved with formation of the inner tube (201, 202). Genes *afp8*, *afp10* and *afp12* produce the needle at the tip of the inner tube (202). The genes *afp2-4* encode for proteins that make up the sheath (198, 202). *afp13* produces tail fibres, these are needed to adhere to target cells prior to injecting a pathogenic package. The *afp14* gene is a tape measure protein that is needed to create particles of the proper size (203). *afp16* is an elongation terminator and forms a dome shape at the top of the particle (201). The putative toxins of the particle are encoded on *afp17-18 (198, 201)*.

The Afp prophage encoded on the pADAP plasmid forms a R-type tailocin apparatus (161, 204). This apparatus consists of a base-plate, outer sheath and an inner tube-like structure with a pointed ending, and tail fibres (194, 205). Under Transmission Electron Microscopy (TEM) it was shown that the particle has two different forms: an extended, and a contracted form where in the latter instance the outer sheath contracts and forcefully protrude the inner tube (161). This system is morphologically similar to other contractile phage-tail derived particles such as type VI-secretion systems (206), R-type pyocins (207) and PVC (191, 201). A simplified figure of Afp mode of action can be seen in Figure 1-10. Full description of the function of each Afp ORF in the formation of the Afp particle can be found in Supplementary Materials Appendix AS.24.

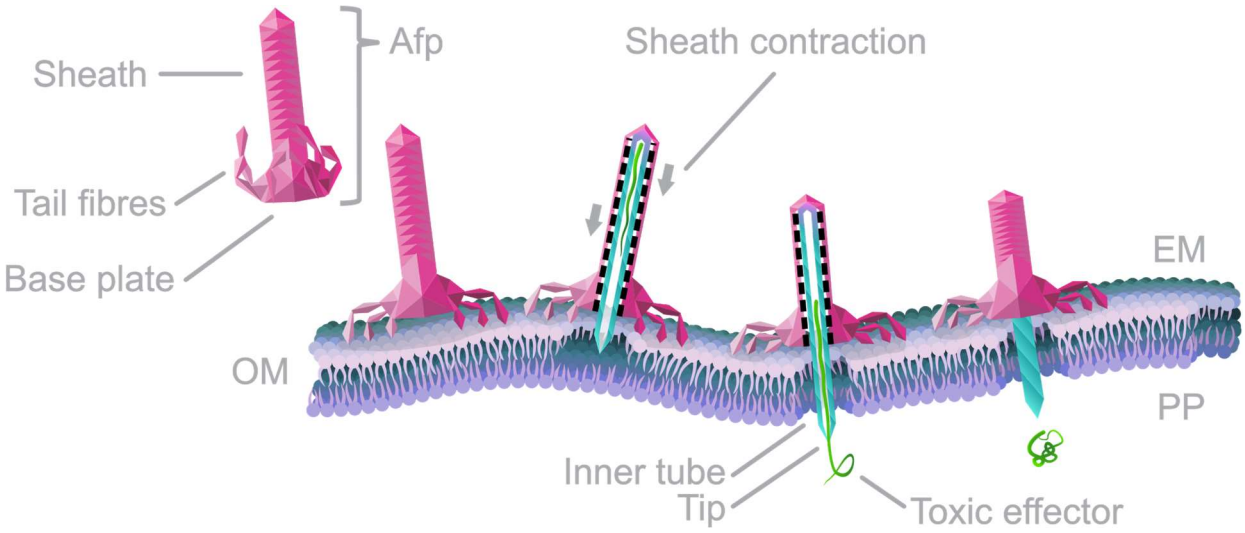


Figure 1-10; A simplified model of the Anti-feeding-prophage particle. This figure shows the outer contractile sheath, the rigid inner tube with attached needle-like end, and the tail fibres. The Afp attaches to the OM, through contraction of the outer sheath, a rigid inner tube is pushed out. This inner tube is then able to release a protein-based toxin load inside the target cell. The means of penetrating the membrane is unknown, as is information about whether the inner tube only breaches the OM or also the IM. Another plausible mechanism is cell mediated endocytosis. This figure is based on the model published by Heymann et al. (194).

It is assumed, based on the function of homologous secretion systems, that the inner tube can penetrate a yet to be defined insect target cell wall and is then able to release toxic effectors into the cytosol. The putative toxic protein effectors are encoded by *afp17* and *afp18* and cause cessation of feeding in the host organism (200). A median dose of approximately 500 Afp particles is required to cause effect in grass grub larvae (201). The function of each individual Afp component is summarized in Supplementary Materials Appendix AS.24.

The Afp is proposed to attach to *C. giveni* larval gut cells, while AGR96X pADAP variant has an Afp orthologue termed AfpX with tail fibres able to attach to both *C. giveni* and *P. festiva* larval gut cells, causing death in both beetle species (69).

The combination of Afp and Sep in the *C. giveni* larvae cause both an anti-feeding phenotype together with voiding of the gut, leading to discoloration of the midgut. This state can persist for up to several months, eventually resulting in the death of the grub by a combination of malnourishment and sepsis due to the eventual invasion of the haemocoel by the bacteria present in the gut (81, 208). Several divergent Sep and Afp clusters were noted in this study and will be further discussed in Chapter 5 and Chapter 6.

1.13 **pADAP variants**

With the discovery of the Afp variant found in AGR96X (69), together with its unique hyper-pathogenic phenotype, it became more apparent that novel pathogenic pADAP variants could be found. These novel variants may be advantageous to the bacterium in particular niches. It has been know that there is some genetic variation in *Serratia* based plasmids (69, 84), but it was hypothesised that pADAP in *S. entomophila* would be clonal, and that HGT of PAIs is what is causing the different phenotypes and size variation of the plasmids. This hypothesis was further investigated by Dodd et al. (94) in 2003. Unfortunately, due to the high costs of sequencing complete bacterial genomes and the minimal bioinformatic resources available at the time, the sequencing data analysed by Dodd et al. (94) was limited. They did however note a Sep cluster being encoded on a plasmid present in *Y. frederiksenii* strain 49 (90), which was confirmed in this study, and showed it to be encoded on a non-pADAP type plasmid. Besides the *Y. frederiksenii* 49 non-pADAP plasmid carrying Sep homologs, several other non-pADAP *Serratia* plasmids encoding for the Sep components were also identified. This study also found two pADAP-type plasmids, similar to ones found in *S. proteamaculans*, residing in *S. liquefaciens* isolates. These and several other major findings will be discussed further in subsequent chapters.

1.14 **Hypothesis and goals**

The research described in this thesis originated from a collaboration within the Bio-Protection Research Centre, involving AgResearch limited, Lincoln University, Massey University and Otago University. Through this collaboration, several projects were created looking into different aspects of insecticidal *Serratia* species, with the major interests being HGT of the pADAP plasmid and pADAP associated elements, chromosomal divergence and its effect on plasmid transmission and stability, and the ecology of the bacteria in the grass grub as well as in the soil. This project and thesis are focussed on the HGT and plasmid-based elements of the *Serratia* research.

Previous DraI Restriction Fragment Length Polymorphism (RFLP) analysis, performed on several *S. entomophila* and *S. proteamaculans* plasmids showed identical band patterns for all *S. entomophila* megaplasmids but highly diverse patterns for those obtained from *S. proteamaculans* isolates (90, 94). This led to the hypothesis that pADAP could potentially be unstable in *S. proteamaculans* cells, or that HGT of virulence islands happens at a higher rate in *S. proteamaculans*, resulting in pADAP plasmids of varying sizes and potentially pADAP plasmids that acquired novel regions in place of the Afp and Sep PAIs.

To look more closely into the correlation between pathotype and genotype of the pADAP plasmid variants, 76 bacterial isolates from the AgResearch strain library were sequenced. The goal was to determine the distribution of pADAP, the distribution of the pathogenic elements of pADAP, and understand why the hyper-pathogenic pADAP-type plasmids as described by Hurst et al. (69), are not more widespread if they carry the more efficient means of affecting grass grubs. These data would allow the determination of whether incompatibility exists between chromosomal hosts and certain pADAP plasmids that prevents proliferation of the pADAP plasmids. The data would also allow for analysis of the distribution of HGT elements and their position on the pADAP plasmids, in order to determine if there is evidence of acquisition or loss of elements, such as genetic islands, as a means of switching pathogenicity on or off.

In the following chapters the vast amounts of plasmid variation will be discussed as well as the phenotypes associated with some of these variations, and what this information tells us about the origin of the plasmid and why there is so much genotypic and phenotypic variation. Each investigated aspect of the pADAP plasmid, it's variants and phenotypical classifications is written in a separate chapter to keep results and discussions condensed and on point.

Chapter 2 Material and methods

2.1 AgResearch bacterial isolate library

Several isolated wildtype strains of *Serratia entomophila*, *S. ficaria*, *S. grimesii, S. liquefaciens*, *S. marcescens*, *S. proteamaculans*, and *Y. frederiksenii* were used in this study (Table 2-1). The isolation and classification of most of these strains have been published in the thesis "Horizontal transfer of plasmid borne insecticidal toxin genes of *Serratia* species" by Dodd (94). Additional isolates, specifically strains that cause atypical phenotypes such as hyper-virulence, have been isolated and classified by Mark Hurst [unpublished data] or were obtained from the AgResearch strain library [unpublished data]. Isolates are described in Supplementary Materials S.1.

2.2 Microbiological methods

2.2.1 Medium

In this project several types of medium were used. Culturing of isolates was undertaken in Lennox L Broth Base (LB-broth, Invitrogen) liquid medium (209), which was prepared according to manufacturers' specifications. For nutritional stress experiments, M9 minimal medium was made according to the recipe in Supplementary Materials S.2. For most experiments ultrapure water was used, this water is double-distilled then filtered using the Milli-Q® Integral Water Purification System.

2.2.2 Laboratory strains

Several commercial and non-commercial bacterial strains, as well as vector systems, were used in this project (Table 2-2).

2.2.3 Mutants and vector constructs

Several mutants and cloned constructs were designed and used in this project (Table 2-3).

2.2.4 Antibiotics and substrates

Antibiotics and other specific additives were used to select for plasmids, inserts or cross validation for absence of either. Antibiotics/additives and their standard concentrations are listed in Supplementary Materials S.5.

Table 2-1: Isolates and species used in this study and their associated pathogenic status.

Genus	Species	Total	Pathotype	Total	Isolate names
Serrata	*proteamaculans*	(50)	Chronic	(5)	4, 142[1], 145, M, RM5[1]
			Hyper	(6)	1129, 10Novel, 1A, 20093, LC, Agr96X[1,2]
			Mixed	(7)	143[1], 1048[1], 12d, CfB, SpF, R10, Sprot5
			Non-path	(32)	1, 149[1], 163, 299, 336, 465, 495, 591, 1071, 1137, 1457, 1769, 1770, 1772, 3041, 12a, 12newD, 25E, 28F, C[4], D, E, F28, G, I, J[3], K, Man4[1], Puna18b, Sm1a, Sm2a, Sprot1[1]
	entomophila	(17)	Chronic	(8)	176, 210, 398, 626, 1100, A1MO2[4], Diarr, Moraki#2
			Non-path	(9)	158, 219, 220, 345, 440, 442, 477, D1, Man3
	liquefaciens	(3)	Non-path	(3)	376, 377, 695
	marcescens	(4)	Non-path	(3)	294, 1573, 1577
			Unknown	(1)	WVU-005[5]
	ficaria	(1)	Non-path	(1)	457
	grimesii	(1)	Non-path	(1)	348
Yersinia	*frederiksenii*	(1)	Non-path	(1)	49

[1] Samples were sequenced both on the Illumina and the PacBio platform.

[2] The partial sequence for the plasmid of Agr96X has been published prior to this study (69).

[3] Sequencing data was insufficient for proper assembly and were ignored in further analysis.

[4] The sequence for the plasmid of A1MO2 was obtained from NCBI, accession number [NC_002523], as it has been published prior to this study (83).

[5] Sequence of plasmid WVU-005-1 and WVU-005-2 was obtained from NCBI, accession number [CP041127] and [CP041128] respectively, sequenced by Ranjbaran et al. [PRJNA545504].

Table 2-2: Strains and vectors used in this study. Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline

Name	Reference	Cassette(s)	Description
Plasmids			
pACYC184	Chang et al.(210)	Cm^R	Chloramphenicol O-acetyltransferase
		Tc^R	Tetracycline efflux MFS transporter
pACYC177	Chang et al.(210)	Km^R	APH(3')-I family aminoglycoside O-phosphotransferase
		Ap^R	Beta-lactamase TEM-1 variant
pHP45Ω	Prentki et al. (211)	Sm^R, Sp^R	Streptomycin 3'-adenylyltransferase
pUCP30T-GFPmut3	Barbier et al. (212)	GFPmut3(213)	Gm^R, Green fluorescent protein mut3.1
pGEM®-T Easy	Promega Corp.		Ap^R, linearized TA cloning vector
pAY2-4	Shaw et al. [190]		Ap^R, pBAD derivative arabinose expression vector, a.k.a as pAVR2, pAVR2-4 and pARA inhouse
pJP5603	Riedel et al. (214)		Kn^R, Suicide vector
pJP5608	Riedel et al. (214)		Tc^R, Suicide vector, no sequence available
143SpRVpJP5603	Hurst [unpublished data]		Construct made for targeted deletion of portion of *sepA* and *sepB*
pJP5603_Scon1	Watson [unpublished data]		Construct made to tag chromosome with Sp^R cassette
E. Coli strains			
Dh10ß	Durfee et al. (215)		
TransforMax™ EC100D™ pir+	Epicentre®		pir+, electrocompetent cells
ST18	Thoma et al. (216)		ΔhemA, requires 5-aminolevulinic acid complementation
AAECO72A	Blomfield et al. (217)		Δfim
Serratia strains			
5.6	Glare et al. (218)		
3041 / Tukino	Glare et al. (218)		
5.6_Chrom_Sp	Watson [unpublished data]		Sp^R tagged 5.6 strain
3041_Chrom_Sp	This study		Sp^R tagged 3041/Tukino strain
A1MO2	Grimont et al. (18)		
XΔNOVa	Hurst [unpublished data]		pAGR96X mutant with v*apBC* complex knocked out

Table 2-3: Cloned and mutated vectors and isolates constructed in this study. Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline

Name	Vector(s)	Cassette(s)	Description
Working strains			
ST18	pJP5608_591_Phage_Head_KO	TcR, $Sm^{R/}Sp^{R}$	Construct of p591 phage head region with SmR/SpR, cloned into pJP5608
ST18	pJP5608_591_Phage_Tail_KO	TcR, Cm^{R}	Construct of p591 phage tail region with CmR, cloned into pJP5608
ST18	pJP5603_BB_Cm	Km^{R}, Cm^{R}	Construct of pADAP PsiA region with CmR, cloned into pJP5603
ST18	pJP5608_BB_Cm	Tc^{R}, Cm^{R}	Construct of pADAP PsiA region with CmR, cloned into pJP5608
ST18	pJP5608_Sep_Spec	Tc^{R}, Sm^{R}/Sp^{R}	Construct of pADAP SepB region with SmR/SpR, cloned into pJP5608
ST18	pJP5608_AFP_Kan	Tc^{R}, Km^{R}	Construct of pADAP Afp2 region with KmR, cloned into pJP5608
ST18	pJP5608_1129_Col_KO	Tc^{R}, Cm^{R}	Construct of p1129 Colicin region with CmR, cloned into pJP5608
ST18	pJP5608_BB_Kan	Tc^{R}, Km^{R}	Construct of pADAP PsiA region with KmR, cloned into pJP5608
ST18	pJP5608_BB_GfpMut3	Tc^{R}	Construct of pADAP PsiA region with GFPMut3, cloned into pJP5608
ST18	pJP5608_149Nov1	Tc^{R}, Cm^{R}	Construct of p149 novel region 1 with CmR, cloned into pJP5608
ST18	pJP5608_149Nov2	Tc^{R}, Sm^{R}/Sp^{R}	Construct of p149 novel region 2 with CmR, cloned into pJP5608
ST18	pJP5608_SEFABC_KO	Tc^{R}, Cm^{R}	Construct of pADAP ΔSefABC region with CmR, cloned into pJP5608
ST18	pJP5608_BB_Amp	Tc^{R}, Ap^{R}	Construct of pADAP PsiA region with ApR, cloned into pJP5608
EC100D	pARA_SEF_CM	Ap^{R}, Cm^{R}	pAY2-4 vector containing SefA-J region in frame with arabinose promotor
3041_Chrom_Spec	-	$Sm^{R/}Sp^{R}$	3041 tagged with SmR/SpR using pJP5603_Scon1
Knockout mutants			
149	p149Δnov1	Cm^{R}	149 with p149Δnovelregion1::Cm^{R}
149	p149Δnov2	$Sm^{R/}Sp^{R}$	149 with p149Δnovelregion2::Sm^{R}/Sp^{R}
149	p149Sprv	Km^{R}	149 with p149ΔSepBC::Km^{R}
591	p591ΔPhage_Tail	Cm^{R}	591 with p591ΔPhageTail::Cm^{R}
591	p591ΔPhage_Head	$Sm^{R/}Sp^{R}$	591 with p591ΔPhageHead::Sm^{R}/Sp^{R}
1129	p1129ΔColicin	Cm^{R}	1129 with p1129ΔColicinregion::Cm^{R}

Expression mutants			
AAE072CA	pARA_SEF_Cm	Ap^R, Cm^R	ΔFIM strain AAE072CA with pAY2-4::SefA-J
AAE072CA	pARA	Ap^R	ΔFIM strain AAE072CA with pARA
A1MO2	pADAP_SEF_ABC_KO	Cm^R	A1MO2 with pADAPΔSefABC
A1MO2	pADAP, pUC30TGFPMut3	Gm^R	A1MO2 with pADAP, GFP
A1MO2	pADAP_SEF_ABC_KO, pUC30TGFPMut3	Cm^R, Gm^R	A1MO2 with pADAPΔSefABC, GFP

Reporter mutants			
142	p142_AFP_Kan	Km^R	Afp2::Km^R
142	p142_BB_Cm	Cm^R	PsiA::Cm^R
142	p142_BB_SEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
142	p142_SEP_Spec	Sm^R/Sp^R	SepB::Sm^R/Sp^R
143	pU143_BB_Cm	Cm^R	PsiA::Cm^R
143	pU143_BBSEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
145	p145_AFP_Kan	Km^R	Afp2::Km^R
145	p145_ALL	Cm^R, Sm^R/Sp^R, Km^R	PsiA::Cm^R, SepB::Sm^R/Sp^R, Afp2::Km^R
145	p145_BB_Cm	Cm^R	PsiA::Cm^R
145	p145_BB_SEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
145	p145_SEP_Spec	Sm^R/Sp^R	SepB::Sm^R/Sp^R
149	p149_BB_Cm	Cm^R	PsiA::Cm^R
149	p149_BB_GFPMut3	-	PsiA::GFPMut3
176	p176_BB_Cm	Cm^R	PsiA::Cm^R
176	p176_BBSEP	CmR, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
176	p176_BBSEP	CmR, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
176	p176_SEP_SPEC	Sm^R/Sp^R	SepB::Sm^R/Sp^R
210	p210_AFP_Kan	Km^R	Afp2::Km^R
210	p210_ALL	Cm^R, Sm^R/Sp^R, Km^R	PsiA::Cm^R, SepB::Sm^R/Sp^R, Afp2::Km^R
210	p210_BB_Cm	Cm^R	PsiA::Cm^R
210	p210_BB_SEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R

28F	p28F_BB	Cm^R	PsiA::Cm^R
299	p299_BB_Cm	Cm^R	PsiA::Cm^R
299	p299_BB_GFPMut3	-	PsiA::GFPMut3
336	p336_BB_Cm	Cm^R	PsiA::Cm^R
345	p345_BB_Cm	Cm^R	PsiA::Cm^R
345	p345_BBSEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::SmR/SpR
376	p376_BB_Cm	Cm^R	PsiA::Cm^R
376	p376_BB_Cm	Cm^R	PsiA::Cm^R
377	p377_BB_Cm	Cm^R	PsiA::Cm^R
398	p398_ALL	Cm^R, Sm^R/Sp^R, Km^R	PsiA::Cm^R, SepB::Sm^R/Sp^R, Afp2::Km^R
398	p398_BB_Cm	Cm^R	PsiA::Cm^R
398	p398_BBSEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
465	p465_BB_Cm	Cm^R	PsiA::Cm^R
465	p465_BB_GFPMut3	-	PsiA::GFPMut3
626	p626_BB_Amp	Ap^R	PsiA::Ap^R
626	p626_BB_Cm	Cm^R	PsiA::Cm^R
626	p626_BB_Cm	Cm^R	PsiA::Cm^R
626	p626_BB_GFPMut3	-	PsiA::GFPMut3
626	p626_BB_Kan	Km^R	Afp2::Km^R
626	p626_BBSEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
1048	p1048_BB_Cm	Cm^R	PsiA::Cm^R
1048	p1048_BB_Cm	Cm^R	PsiA::Cm^R
1048	p1048_BB_CM	Cm^R	PsiA::Cm^R
1048	p1048_BB_GFPMut3	-	PsiA::GFPMut3
1048	p1048_BB_GFPMut3	-	PsiA::GFPMut3
1071	p1071_BB_CM	Cm^R	PsiA::Cm^R
1100	p1100_AFP_Kan	Km^R	Afp2::Km^R
1100	p1100_ALL	Cm^R, Sm^R/Sp^R, Km^R	PsiA::Cm^R, SepB::Sm^R/Sp^R, Afp2::Km^R
1100	p1100_BB_Cm	Cm^R	PsiA::Cm^R
1100	p1100_BB_SEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R

1129	p1129_AFP_Kan	Km^R	Afp2::Km^R
1129	p1129_BB_Cm	Cm^R	PsiA::Cm^R
1129	p1129_BB_GFPMut3	-	PsiA::GFPMut3
1129	p1129_BBSEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
1137	p1137_BB_Cm	Cm^R	PsiA::Cm^R
1137	p1137_BB_Cm	Cm^R	PsiA::Cm^R
1137	p1137_BB_GFPMut3	-	PsiA::GFPMut3
1769	p1769_BB_Cm	Cm^R	PsiA::Cm^R
1769	p1769_BB_GFPMut3	-	PsiA::GFPMut3
1769	p1769_BBSEP	Cm^R, Sm^R/Sp^R	PsiA::Cm^R, SepB::Sm^R/Sp^R
A1MO2	pA1MO2_AFP_KAN	Km^R	Afp2::Km^R
A1MO2	pA1MO2_SEP_SPEC	Sm^R/Sp^R	SepB::Sm^R/Sp^R
D	pD_BB_Cm	Cm^R	PsiA::Cm^R
D	pD_BB_GFPMut3	-	PsiA::GFPMut3
MHX	pMHX_BB	Cm^R	PsiA::Cm^R
MHX	pMHX_BB_Amp	Ap^R	PsiA::Ap^R
MHX	pMHX_BB_GFPMut3	-	PsiA::GFPMut3
MHX	pMHX_BB_Kan	Km^R	Afp2::Km^R
pIDIA	pIDIA_AFP_KAN	Km^R	Afp2::Km^R
pIDIA	pIDIA_BB_CM	Cm^R	PsiA::Cm^R
pIDIA	pIDIA_SEP_SPEC	Sm^R/Sp^R	SepB::Sm^R/Sp^R
Puna18b	pPun18b_BB_Cm	Cm^R	PsiA::Cm^R
Puna18b	pPuna18b_BB_GFPMut3	-	PsiA::GFPMut3
RM5	pRM5_BB	Cm^R	PsiA::Cm^R

BB: region between p*siA* and *sea40*, Afp: region between *afp3* and *afp4*, Sep: region between a hypothetical ORF and *sepC*, for further details see Section 2.2.14.

2.2.5 Preparation of plasmid DNA

Cells were grown overnight (o/n) in 28 ml McCartney bottles containing 3 ml of LB-broth, and incubated on a Ratek orbital incubator, at 30C°, shaking at 200rpm o/n, for 16-18h. One mL of these cultures was spun down in an Eppendorf 5424 benchtop centrifuge, at 6000 g, the supernatant aspirated and the pellets washed two times with sterile UPW. Plasmid DNA extraction was performed using the High Pure Plasmid Isolation Kit (Roche), according to the manufacturers' specification.

2.2.6 Polymerase chain reaction (PCR)

PCRs were performed on either a Bio-Rad C1000 Touch™, an Eppendorf Mastercycler EP Gradient S or a Bio-Rad MyCycler™ Thermal Cycler. PCRs were run using the reagents and program settings as shown in Supplementary Materials S.4, unless stated otherwise. PCR products were purified using the High Pure PCR Product Purification Kit (Roche).

2.2.7 Gel Electrophoresis (GE) assessment

Three µl of DNA sample were mixed with 1 µl of DNA Gel Loading Dye (Thermo Scientific™), and loaded on a 1% agarose gel, containing 0.1% RedSafe™ Nucleic Acid Staining Solution (iNtRON biotechnology). Three µl of Quick-Load® 1 kb DNA Ladder (New England Biolabs© Ltd.) was added as reference to determine band size. The gel was run for 60 min using a Bio-Rad Mini-Sub Cell GT Horizontal Electrophoresis System set to 100 V. Gels were assessed using a Uvitec Uvidoc HD2 20MX gel documentation system allowing smear patterns to indicate subpar PCR and for proper size of bands to indicate proper primer functionality to be observed.

2.2.8 Sequence validation

DNA was tested for quality as described in Section 2.2.7. Roughly 50 µl of samples, containing ≥25 ng/ml DNA, were of good quality were sealed and sent to Macrogen Korea (South Korea) for their Standard-seq sequencing service. Resultant sequence was aligned to the relevant sequence assembly allowing sequence validation.

2.2.9 Restriction digestion

DNA was digested using the appropriate reagents as outlined in Supplementary Materials S.7. Samples were incubated at 37°C for 1.5 hours. Subsequently, DNA was purified by adding 10 µl of 3M Sodium Acetate (NaAc), followed by adding 200 µl of >95% ethanol, and mixing briefly. This was incubated at room temperature for ~2 min. Sample was pelleted down at 16000 g for 5 min. Supernatant was aspirated and sample was dried at 37°C for ~5 min. The pellet was resuspended in 7 µl of UPW.

2.2.10 Megaplasmid visualization

The Kado and Liu megaplasmid visualization protocol (219) was used to visualize the presence of >10 Kb plasmids. The recipe for buffers used for this protocol can be found in Supplementary Materials S.3. *Serratia* cells were grown 16-18 h o/n at 30°C, shaking at 200 rpm. Of the o/n culture, 300 µl was spun down at 16000 g for 1 min. Supernatant was aspirated and the pellet was resuspended in 100 µl of E buffer. To the solution, 200 µl of lysis solution was added, and samples were slowly inverted to mix. Samples were then incubated at 55°C for 60 minutes in an Eppendorf thermomixer Comfort. Phenol/chloroform extraction was performed by adding 300 µl of Phenol:Chloroform:Isoamyl Alcohol 25:24:1 saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma-Aldrich) solution to the sample, and slowly inverting for 2 min. The top aqueous phase was transferred to a clean tube for further use. Thirty µl of sample was mixed with 6 µl of DNA Gel Loading Dye (Thermo Scientific™) and loaded on a 0.7% agarose gel. A lambda dna HindIII Marker was used as a ladder. The gel was run for 4 h in a BRL Horizon™ 11-14 Midi Gel System, with a C.B.S. Scientific company, inc. EPS-300 IIV power supply set to 80 V. The gel was then strained in Ethidium Bromide for 5 min and assessed using a Uvitec Uvidoc HD2 20MX gel documentation system.

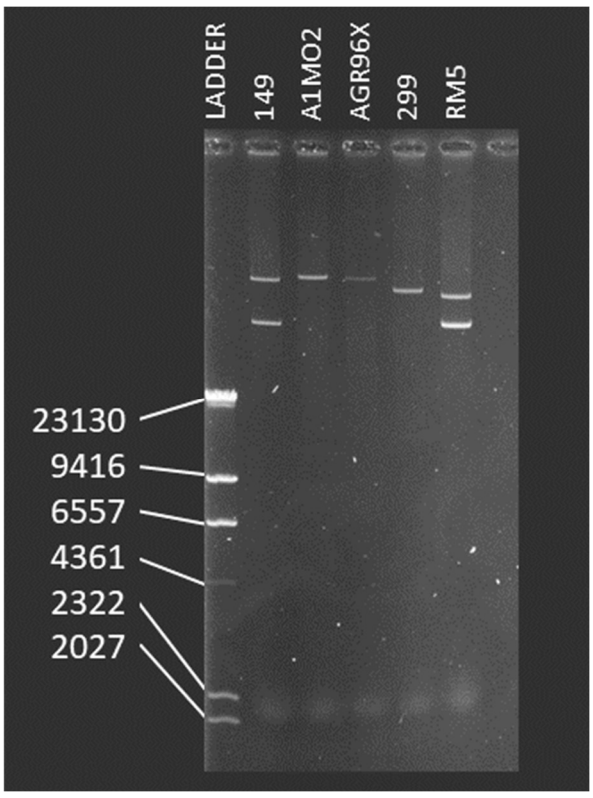


Figure 2-1: Example of Kado and Liu (219) megaplasmid visualization gel. The lambda DNA HindIII Marker is shown on the left, the isolate names on top. Several isolates contain two or more plasmids as show by isolate 149 and RM5.

2.2.11 Fusion PCR

Mutants were made using a two-step fusion PCR approach as described by Szewczyk et al. (220, 221). Plasmid templates were obtained from cultures as described in Section 2.2.5. A PCR reaction, as outlined in Section 2.2.6, was performed on the extracted plasmid template to amplify approximately 500 bp region upstream and downstream of the insertion site, as seen in step A in Figure 2-2. Primer set P1-P3 was used to amplify up-fragments, where P3 is a ~40 bp primer of which half overlaps with the cassette that is meant to be fused. Primer set P4-P6 was used to amplify down-fragments, where, like P3, P4 consisted of ~40 bp of which half overlapped with the cassette. The cassette was amplified from its host vector using primer set CF-CR.

The fusion PCR was performed as described in Section 2.2.6, but instead of 1 µl template, the purified up-fragment, down-fragment and the cassette amplicons were used as template in a 1:1:1 µl ratio, and 37 µl UPW to compensate for the increased template volume. Primers P2 and P5 were used for the fusion; they are nested primers approximately 50 bp inwards of the up- and down-fragments, these primers contain three adenine nucleotides and a SacI restriction enzyme (RE) site on the 5' side of the primer.

The quality of fused amplicon construct was tested as described in 2.2.7. A schematic of the two-step fusion PCR can be seen in Figure 2-2. All primers used for the construction of these fused amplicons can be found in Supplementary Materials S.6.

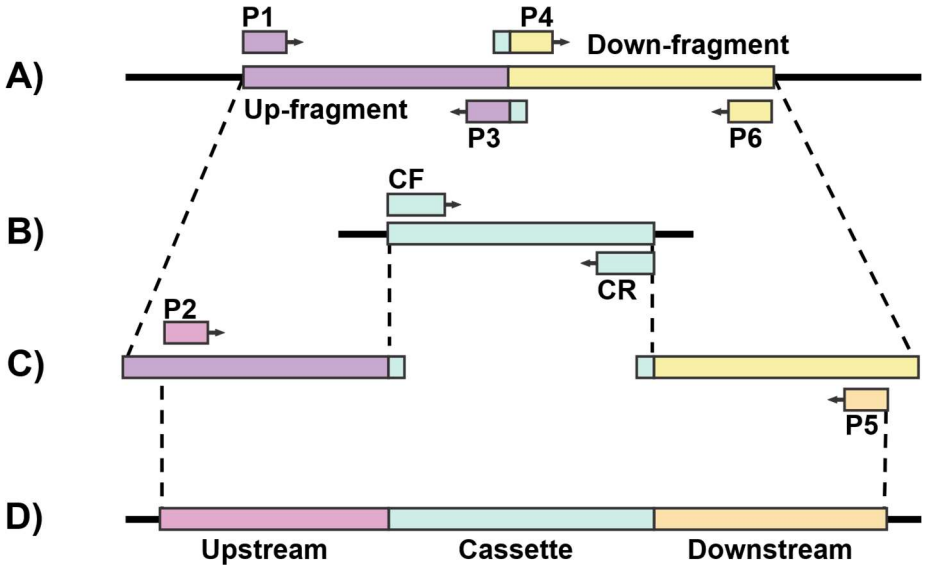


Figure 2-2: A schematic of the two-step fusion PCR. A) a ~500 bp region up- and downstream of the insertion site were amplified. B) a cassette of choice is amplified, these are often ~1000 bp. C) All PCR products are purified and added together as template in a new reaction at a 1:1:1 µl ratio. D) a ~2000 bp fused amplicon should be formed containing a fused amplicon of the upstream region, cassette and downstream region.

2.2.12 Cloning

Fused PCR constructs obtained using the method described in Section 2.2.11, were ligated into the pGEM®-T Easy cloning vector (Promega Corp.) according to suppliers' protocol. This commercial linearized vector has T overhangs, and because of the added poly A tails to the fused PCR constructs, ligation into the vector is efficient. Samples were left to ligate o/n at 4°C. The ligate was electroporated into TransforMax™ EC100D™ pir+ electrocompetent *E. coli* strains in a 2:40 µl ligate to cell ratio. Electroporation was undertaken using a Bio-rad *E. coli* Pulser™, set to 2.5 KV, according to the manufacturer's protocol. Electroporated cells were then resuspended in LB-broth and incubated at 37°C for 1 h in a Contherm Digital Five Series incubator. Cells were pelleted at 6000 g for 3 min, supernatant aspirated and the pellet resuspended in 100 µl of LB-broth. Cells were then plated out on LB-agar plates with the appropriate antibiotics and incubated at 37°C. The plates were supplemented with X-gal50, as the pGEM allows for blue–white screening. White colonies were patched to fresh plates. Healthy colonies were cultured o/n and miniprepped as described in Section 2.2.5 and screened with a restriction digestion as outlined in Section 2.2.9 and sequencing approach as described in Section 2.2.8.

Positive vector mutants were miniprepped as outlined in Section 2.2.5, and digested using SacI REs (New England Biolabs© Ltd.) as described in Section 2.2.9. This digestion releases the fragments from the pGEM vector. The pJP5608 suicide vector (214) was also digested using SacI to open up the vector. A ligation of both products was performed using T4 DNA ligase (Thermo Scientific™) in order to ligate the insert into the suicide vector. Ligated products were electroporated into the ST18 *E. coli* strain (216). ST18 is a ΔhemA strain and therefore needs to be supplemented with 5-aminolevulinic acid (ALA) in order to grow. ST18 colonies were again validated using restriction digestion (Section 2.2.9) and a sequencing approach (Section 2.2.8). Unless stated otherwise, all constructs were made this way. The final constructs can be found in Supplementary Materials S.8.

2.2.13 Conjugation

Insertion of a tag cassette or knockout of genes was undertaken using the appropriate construct made as described in Section 2.2.11 present in a suicide plasmid in *E.coli* strain ST18 as described at Section 2.2.12. The appropriate ST18 donor strain and *Serratia* recipient were independently grown o/n at their respective temperatures in LB-broth containing the appropriate antibiotics, and for ST18, ALA50. One ml of each o/n culture was pelleted at 6000 g and the resultant pellets washed twice in fresh LB-broth. The pellets were then resuspended in 500 µl of LB-broth and 50 µl of both the recipient and the donor cultures were added to a LB-agar plate containing ALA50. The 50 µl aliquots were mixed manually by gently rotating the plate, and the plate was incubated at 30°C in a Barnstead Lab-Line L-C incubator for 6-8 hours.

Subsequently, 1 ml of fresh LB-broth was added to the plate, and the grown colonies were scraped off the top of the LB-agar with a sterile spreader and resuspended in the LB-broth. Dilutions of the resuspended conjugants were then plated out on LB-agar plates containing the appropriate antibiotics and incubated o/n at 30°C. Colonies were patched the next day on plates selecting for insert and counter selecting for the suicide vector. Colonies were screened by miniprepping them as described in Section 2.2.5 and sequence validated as described in Section 2.2.8. Validated mutants were transferred to TS/80-MX Protect Microorganism Preservation System vials (Technical Service Consultants Ltd.) and stored in a Thermo Scientific TSX60086V Ultra-Low Freezer (-80°C).

2.2.14 Tagged pADAP regions

Three highly homologous sequence regions were picked, in which to insert antibiotic resistance cassettes that would allow tracking of plasmid presence, and presence of pathogenicity associated islands. For the backbone, a highly conserved region between *psiA* and *sea40* was selected (Figure 2-3). For tagging of the Sep PAI a region between a hypothetical protein (*orf4*) and *sepC* was chosen. For the Afp virulence encoding region, a highly conserved intergenic region, located between *afp2* and *afp3* was picked. Recombination was undertaken as outlined in Section 2.2.13. A full list of tagged plasmids can be found in Section 2.2.3.

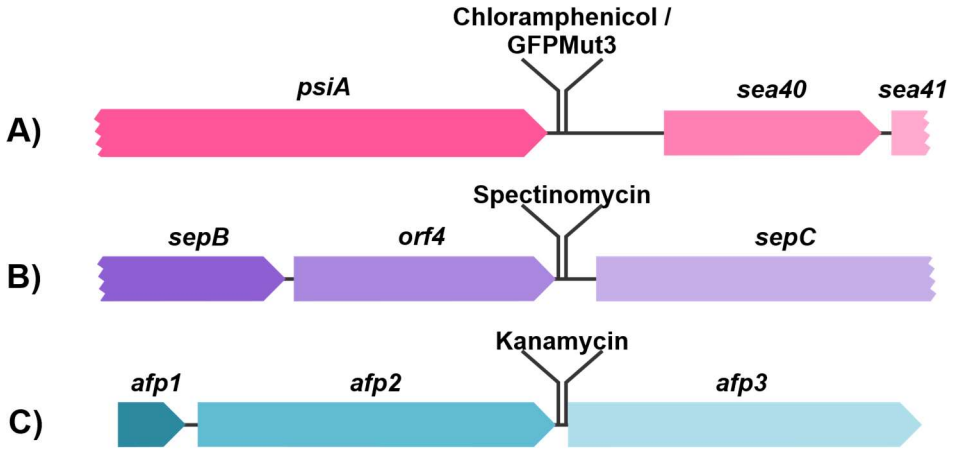


Figure 2-3: A schematic of the three primary locations for pADAP plasmid tagging, and the antibiotic cassettes used. The three main locations for plasmid tagging were A) The backbone, in an intergenic region with no predicted function between *psiA* and *sea40* which is a mRNA interferase as described in Supplementary Materials S.8, B) The Sep region, in between a hypothetical protein with a domain of unknown function (*orf4*) and the *sepC* ORF, C) the Afp region, in an intergenic region between *afp2* and *afp3*.

All steps in creating these tagged plasmids is validated using PCR and Sequencing. And example of the PCR workflow can be found in Figure 2-4.

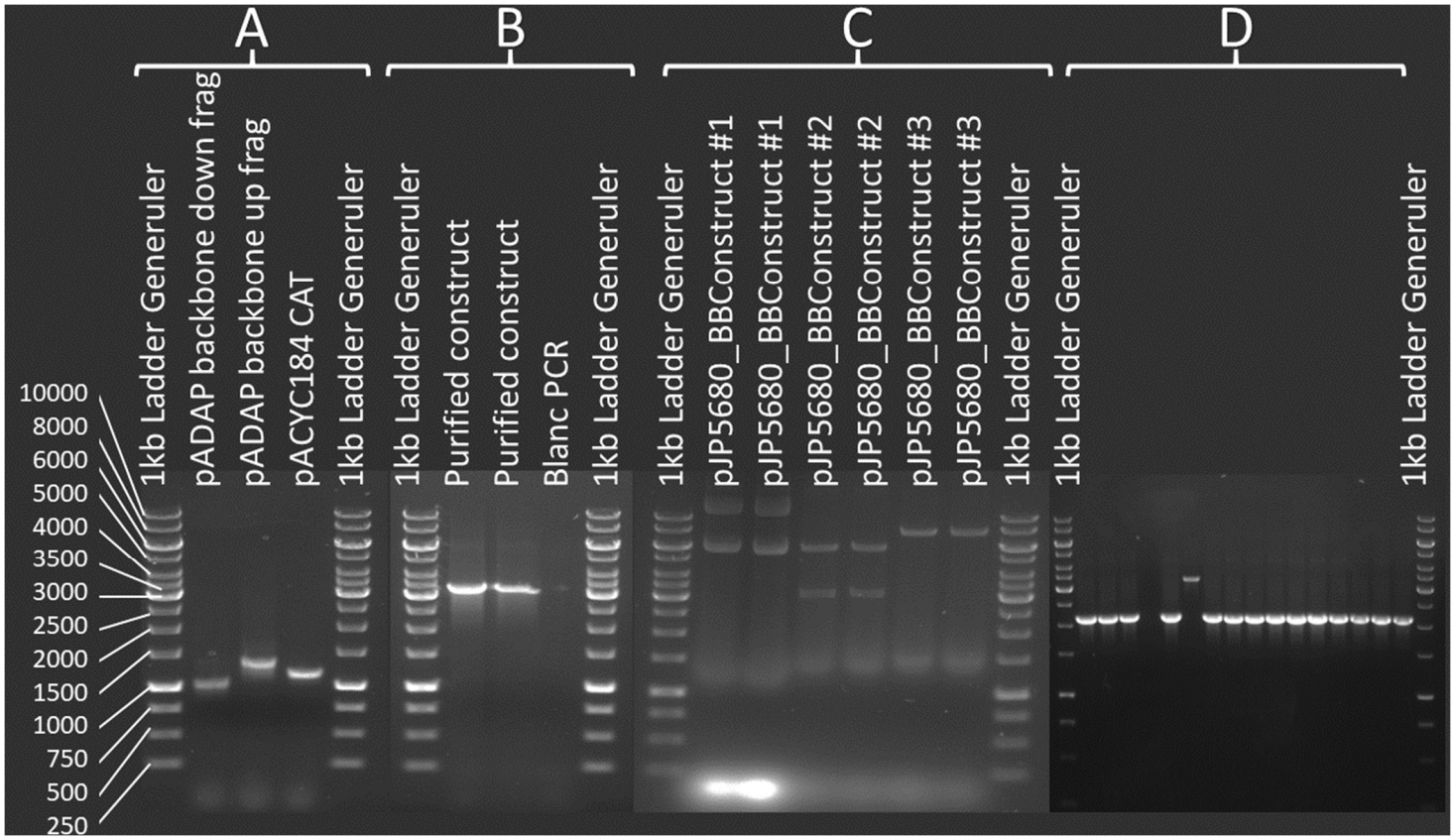


Figure 2-4: Example of PCR validation workflow. This example depicts validation of backbone (BB) construct containing a Cm^R cassette (CAT) obtained from pACYC184. A) Primer pairs were made for regions flanking targeted insertion site (P1-P3, P4-P6), as well as for the antibiotic cassette of choice, amplicons were checked for size. B) Primer pair, nested ~50 bp inside of the flanking fragments (P2-P5), were used to fuse the amplicons into a construct. C) This construct was ligated into a vector of choice, and transformed into a cell, the following day colonies were tested by miniprepping the vector and doing one or several restriction digestions. In this case we expected two band, one for insert (3 Kb) as well as one for the pJP5680 vector (6.7 Kb), making vector #2 the only valid one. D) vectors were conjugated into target isolate, allowing for double recombination. Potential double recombinants were picked and a colony PCR is performed the P1-P6 primer pair. Successful recombinants should generate a larger amplicon than WT (as seen in lane 7). The isolate is miniprepped and another PCR using primer pair P1-P6 was performed. The amplicon was then sent for sequencing.

2.2.15 HGT growth experiments

Isolates were grown in 28 ml McCartney bottles containing 3 ml of growth medium without antibiotics. Samples were cultured on a Ratek orbital incubator, at 30C°, shaking at 200 rpm o/n. Each successive day 30 µl of culture was sub-cultured in a new McCartney bottle containing 3 ml of fresh medium without antibiotics. At the tenth day, dilutions of the culture were plated out on LB-agar plates and incubated at 30C° for approximately 24 h.

2.2.16 Plasmid stability flow cytometry experiments

Samples were cultured in triplicate as outlined in Section 2.2.15. At day ten, 0.5 ml of sample were washed twice by spinning it down at 6000g for 3 min and pellet resuspended in 1 ml of 25% glycerol. Gibco® Phosphate-Buffered Saline (PBS) Tablets were used the manufacturer's protocol to make PBS stock buffer. Samples were prepared for flow cytometry by filling 5 ml round bottom Falcon™ Test Tubes with 2 ml of PBS stock buffer, and 0.4 µl of sample. Samples were analysed using a BD FACSCanto™ II cytometer. Per sample, 10 000 events were recorded using forward scatter (FSC) detector (488/10 nm) to determine the size and shape of the event, side scatter (SSC) detector (488/10 nm) to determine the makeup of the event, and a Fluorescein isothiocyanate (FITC) detector (530/30nm) to detect whether the event is emitting fluorescence in the expected range of GFPMut3 (Ex: 510/9 nm, Em: 530/15 nm) (213).

The wildtype isolate 626 was used as a negative control to determine SSC and FSC values similar to the cells, and 626, bearing a GFPMut3 backbone tagged pADAP mutant, was used to confirm the SSC and FSC values and to determine expected FITC values. Using these values, gates were placed around events that were most likely single copy cells, and samples were analysed using these default gates. Gates had to be redirected slightly for each growth condition as it appears cells had a slight size and content difference in each growth condition based on FSC and SSC values. Fluorescent cells per population was scored using FlowJo® software v10.6.1 (222). The raw data and summary data can be found in Supplementary Materials Appendix AS.33.

2.2.17 Growth curve experiment

The pADAP bearing A1MO2 type-strain (77), and the heat cured 5.6 mutant (81) were cultured o/n in 3 ml LB-broth. A separate Greiner CELLSTAR® 96 well plates was used for each of the three culture conditions; LB, M9 minimal medium and LB-broth spiked with 0.5 mg / ml MitC. The dilution scheme for the plates can be found in Supplementary Materials S.34. Plates were analysed using a BMG Labtech SPECTROstar Nano plate reader with closed lids to prevent spillages. Samples were shaken at 200 rpm at 30°C for 24 h and were scanned at 600 nm every 5 minutes. Minimal condensation on the lid was observed at the 24 h point, but a test read with condensation and without condensation showed this effect was negligible. Data were converted into Excel format using the SPECTROstar Nano MARS V2.10 data analysis software (223).

2.2.18 Phage and bacteriocin purification

Purification of bacteriophage and bacteriocin particles for TEM was performed as described in the protocol by Hockett et al. (224).

2.2.19 Transmission electron microscopy.

Electron Microscopy sciences EMS200-Cu Plastic-coated 200-mesh copper grids were coated with a 3 nm layer of carbon for better heat dispersal followed by glow discharge treatment to make the surface hydrophilic and negatively charged. Three µl of sample was added to the grid and left for 60 s upon which residual liquid was removed using Whatman #1 filter paper. Samples are then negatively stained using 3 µl of 0.7% uranyl acetate (UA) for 45 s. Excess stain was again removed using the Whatman #1 filter paper and the grid left to dry for at least 30 min. Grids were examined in a Morgagni 268D transmission electron microscope (TEM), images were captured using an Olympus Megapixel III digital camera at varying magnifications.

2.2.20 Arabinose induction

Cells transformed with the pAY2-4 or pARA_Sef vector, were grown in 40% LB broth to reduce the amount of the glucose inhibitor and supplemented with 200 µg/ml arabinose for induction. Cells were shaken at 40 rpm at room temperature.

2.2.21 SDS-PAGE

Isolates were induced as described in Section 2.2.20. At specified time points, the cell pellet and culture supernatant were assessed using Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE). 10% SDS-PAGE gels were ran at 200 V for 50 min. Gels were stained using the “Short silver nitrate staining” protocol as described by Chevallet et al. (225). Gels were assessed on a light box, and images taken, which were converted to grayscale for convenience of the viewer.

2.2.22 Sef adherence to carrot

To determine if Sef is associated with adherence to a food source, A1MO2 WT as well as an A1MO2 carrying the pADAP*ΔsefA-C* were transformed with pUCP30T-GFPmut3 (212) to introduce fluorescence of cells. A 5 µl drop of undiluted A1MO2 WT and A1MO2 carrying the pADAP*ΔsefA-C* was pipetted onto separate ~5 mm thick carrot slices, as this is a preferred food type for grass grubs in bioassays. These inoculated carrot slices were left overnight at 15°C with a dampened tissue to prevent dehydration. This was undertaken to ensure potential adherence would occur. The next day the carrot slices were analysed under the microscope (Section 2.2.23), and then transferred to a cup of UPW, to soak. The cup was left to shake slowly at 40 rpm on a Bellco Glass, Inc. Orbital Shaker. After 1 h of soaking, the slices were removed from the cup and rinsed off with UPW and analysed again under the microscope.

2.2.23 Microscopy

Samples were analysed on a Differential interference contrast (DIC) Olympus BX53 microscope. The FitC channel was used to filter the emitted and excited light on the right wavelength for GFPmut3 carrying samples. Emitted light for GFPmut3 mutants was produced with a CoolLED pE-300. Both DIC and FitC Images were recorded with an Olympus DP74 fluorescent digital camera. $Magnification_{System} = Magnification_{Objective} * Magnification_{Eyepiece}$, which results in 10x objective * 10x camera/eyepiece = 100x magnification. Fluorescent images were taken with at 100x magnification at 5 ms exposure unless stated otherwise, DIC images were taken at 100x magnification with exposure time being automatically determined by the software.

2.2.24 Microinjection of bacteria into grass grub larvae

To validate Sef adherence to the grass grub gut, bacterial culture had to be injected directly into the gut prior to feeding, in order to eliminate the food as an adherence factor for the Sef/bacteria. An in-house microinjector apparatus was used to inject approximately three pulses of 0.1 ml of 1:10 diluted culture directly into the oesophagus of the grub. This was performed using a blunted needle to prevent perforation of the foregut lining. Post injection, the grass grub larvae were placed back in clean trays and left overnight at 15°C to recover. The following day the gut was removed from each larva and opened, enabling it to be assessed through microscopy (Section 2.2.23).

2.3 Bioinformatic methods

2.3.1 Preparation of genomic DNA for sequencing

For the purposes of this project, 76 isolates (eventually generating a total of 84 plasmid sequences) were grown o/n in 28 ml McCartney bottles containing 3 ml of LB-broth, and incubated on a Ratek orbital incubator, at 30°C, shaking at 200 rpm o/n, for 16-18 h. One mL of these cultures was pelleted at 6000 g for 3 min, the supernatant aspirated and the cell pellets washed two times with sterile UPW. Genomic DNA extraction was performed using the Bioline ISOLATE II Genomic DNA Kit according to manufacturer's protocol. Sample quality and concentration were determined using a NanoDrop2000 and regular 1% Agarose Gel Electrophoresis (AGE) (226). Samples that passed the quality checks were sequenced at Macrogen Korea (South Korea). Samples were either sequenced on the Hiseq2500 100bpPE platform, using the Truseq nano DNA kit (350 bp insert) or were sequenced using long-read PacBio sequencing. For PacBio sequencing, the PacBio RSII system was used on SMRT cells, using the 10 Kb SMRTbell template library kit for library preparation. For more detailed information about the individual isolates, the method of sequencing, and the library kit used for sample preparation, please refer to Supplementary Materials S.1.

2.3.2 Server setup

For high performance computing the AgResearch computing server was used. This server is setup as described in Table 2-4 and had an additional computer farm for batch or parallel processing that could be accessed through the HTCondor distributed computing software package (227)

Table 2-4: AgResearch main server setup.

Main server setup	
OS:	CentOS Linux release 7.6.1810 (Core)
Architecture:	x86_64
CPU op-mode(s):	32-bit, 64-bit
CPU(s):	24
Thread(s) per core:	2
CPU Model name:	Intel(R) Xeon(R) CPU E5-2643 v4 @ 3.40GHz

2.3.3 Assembling Illumina data

Sequencing data, received in FASTQ format (228), was trimmed using Trim Galore v0.6.1 (229) with a Phred quality score (230) cut-off of 30 (this equates to a base call accuracy of 99.9% or higher). Trim Galore is a wrapper script that uses Cut Adapt v2.10 for trimming (231) and FastQC v0.11.9 (232) for analysing of trimmed read quality. A small test was performed on an Illumina dataset, sequenced before the start of this project, to determine the optimal software and settings for the assembly pipeline.

These assemblers included: SOAPdenovo v2.04-r241 (233), Velveth v1.2.10 (234), SPAdes v3.14.1 (235), PlasmidSPAdes (236) and A5-miseq v20160825 (237). A5-miseq proved to require the least amount of user input, and for most strains in the test set, was able to construct contigs of similar sizes to the estimated plasmid sizes determined by Dodd (94) and Hurst [unpublished data], making it ideal for batch automation. Once assembled, SSPACE v3.0 (238) was used to scaffold any remaining potentially scattered plasmid contigs, and GapFiller v1-10 (239) was used to resolve any unambiguous bases left after scaffolding. The batch scripts written to automate the assembly process can be found in Supplementary Materials

2.3.4 Assembling PacBio data

PacBio data produced by the PacBio RSII system was delivered in three different formats, bas.h5[1], FASTA (240) and FASTQ (228). Per strain 3 SMRT Cells were used to generate sequence reads, resulting in three separate read files per strain. The three FastQ files were concatenated into one and assembled using Canu v2.0 (241). The error rate for a single continuous long read (CLR) is predicted to be around ~14% (242). Although Canu has a built in error correction based on consensus of the assembled reads, a separate base correction tool was applied called PILON v1.23 (243) which uses Illumina reads to correct erroneous base calls. Due to the mechanism behind PacBio sequencing, circular DNA molecules like plasmids often result Circular Consensus Reads (CCS) or contigs that have large overlapping regions at either end of the sequence. Manually checking for similar regions at the ends of a contig and cleaving them manually can result in bad assemblies due to the erroneous assumptions about PacBio and the assembly process. These ends are often created due a single read with several miscalled bases, making it appear as a novel region for the assembler thus extending the contig with a sequence that seems like a repeat of the other end. To resolve this issue a tool called Circlator v1.5.5 was used to remove overhangs in assemblies of circular DNA molecules (244).

2.3.5 Extracting plasmid related contigs

Upon sequencing, the plasmid contigs were extracted manually by taking contigs, matching the estimated sizes described by Dodd et al. (94), looking for relevant plasmid related BlastN (245) hits, and also manually looking for pADAP characteristics on smaller contigs. Four strains had low quality sequencing data that did not yield any workable assemblies either due to contamination, low N50 or not finding appropriate plasmid contigs, leaving 72 workable genomes. The strains that were rejected from further analysis are highlighted in blue in the table in Supplementary Materials S.1.

[1] https://www.ncbi.nlm.nih.gov/core/assets/sra/files/bas.h5ReferenceGuide.pdf

2.3.6 Validating assemblies

For several plasmid contigs, a primers pair was developed, each primer designed ~500 bp downstream of the 5' and ~500 bp upstream of the '3 side of the contig and extending outward, listed in Supplementary Materials S.6. The PCR amplicon should therefore yield an amplicon of at least 1 Kb in length. PCR amplicon validation was performed as outlined in Section 2.2.8. This validation was undertaken for a subset of the sequences at the start of the project. No erroneous plasmid contigs were found using this method therefore it was assumed all further plasmids conformed to the same quality.

2.3.7 Determining plasmid copy using coverage data

This study sequenced nine isolates using both the Illumina and PacBio platform. Of these nine isolates, six contained a pADAP plasmid. The Bowtie short read alignment tool (246), was used to align the illumina reads of the six isolates to their respective PacBio based genome assembly. The Sequence Alignment/Map (SAM) file produced by Bowtie was sorted and indexed using Samtools v1.10 (247) and a faidx index was created for further processing. Scaffold headers and their respective sizes were extracted using the awk text processor commonly available in most Linux based systems. The SAM file is supplied to genomeCoverageBed, which is part of the BEDTools suite v2.29.2 (248), and using the headers extracted with awk, genomeCoverageBed is able to produce a coverage histogram. This histogram shows the number of nucleotides that have x amount of coverage. This output is then processed by a custom R script. Both the bash script used to batch generate the data for all six isolates, as well as the corresponding R script can be found in Supplementary Materials Appendix AS.12.

2.3.8 Annotating plasmid sequences

Annotation of plasmid contigs was performed using Prokka v1.13 (249). Prokka is a rapid annotation pipeline created for prokaryotic genome assemblies. It uses Blast+ (245) to map RefSeq (250) and UniProt (251) proteins to the assembly, it uses HMMR (252) to run profile databases like TIGRFAM (253) and PFAM (254) against the assembly and it uses Prodigal (255) to predict *de novo* CDSs. The non-redundant (NR) bacterial protein dataset from NCBI was added as a user supplemented evidence dataset. Manual evaluation between fully annotated pADAP [NC_002523] obtained from NCBI and the de novo Prokka annotation performed on the pADAP [NC_002523] raw FASTA sequence obtained from NCBI, shows there to be only two missing gene annotations, *hol1*, two genes upstream of *afp1* and *int1* downstream of *afp18*, these have been manually annotated in all pADAP-type plasmids afterwards using Geneious v8.1.5 (256).

2.3.9 Phylogenetic tree construction of RepA orthologs

The RepA AA sequence from pADAP [NC_002523] reference sequence was aligned to the RefSeq non-redundant protein database (nr) (250) of NCBI (257) using the BlastX webservice (245). The top twenty non-redundant hits with valid species names were obtained. A global alignment was made using Clustal Omega v1.2.4 using default settings (258). PhyML v3.1 (259) was used using the Hasegawa-Kishino-Yano (HKY85) nucleotide substitution model (260), to generate 100 bootstrapped trees. A maximum likelihood (ML) tree was generated from these 100 bootstrapped trees by PhyML. The tree was converted into a circular style, and branches were transformed to a cladogram layout. Numbers on the branches represent substitutions per site numbers. The tree branches and labels were coloured to signify respective host range of associated bacteria using CorelDraw X8 (261).

2.3.10 Distance matrices

Regions of interest were aligned using Clustal Omega. Distance matrices were calculated using Geneious (256) Supplementary Materials Appendix AS.18. Matrix was converted into a heatmap using a custom R script, written in R Studio v1.1.463 (262), running on R v3.5.3 (263). The script can be found in Supplementary Materials Appendix AS.19

2.3.11 Sequence comparison using BRIGS

Significant sequence differences between the assemblies of novel pADAP-type plasmids and the reference sequence of pADAP were assessed by utilizing a tool called BLAST Ring Image Generator (BRIG) v0.95 (264). The software takes a reference sequence, in this case the pADAP [NC_002523] sequence obtained from NCBI [185], and uses Blast+ (245) to map query sequences against it. The analysis was performed using all plasmids to determine presence of pADAP elements. The analysis was then repeated using only the plasmids with pADAP elements (Figure 4-5).

2.3.12 Comparing plasmid annotations

Comparison between the large number of annotated plasmids was undertaken using Roary v3.11.2 (265), a BlastP “all-against-all” algorithm (266) that uses identify-match and e-value cut-off setting to group orthologous genes together. A 90% amino acid identity cut-off was used for clustering. The output of Roary is a Comma-separated values file (CSV) and was processed using a custom R script, written in R Studio 1.1.463 (262) running on R 3.5.3 (263). The script can be found in Supplementary Materials Appendix AS.37.

2.3.13 Scoary analysis on Roary output

Further analysis of the Roary output was undertaken using Scoary v1.6.16 (267). Scoary is a scoring tool that analyses absence and presence of features and correlates them to a table of user supplied features. This table contains a column with strain names, and one or several columns containing binary features. The features used here were the species types to determine any outstanding genes associated with specific species, pathotypes divided in Chronic, Hyper, Mixed and Non-Path and the last category was a binary measure for absence or presence of secondary plasmids to see if there is a common gene associated with pADAP plasmids that share their host with additional plasmids. SCOARY was run with the permutation (-e) option set to 50 fin order to generate an empirical p-value for each found gene. The trait table as well as all the outputs can be found in Supplementary Materials Appendix AS.39.

2.3.14 Codon and nucleotide bias analysis

Codon usage was analysed using GCUA v1.0 (268). FASTA files containing all the CDSs per genome were imported into GCUA to process them and output their corresponding codon usage tables. Nucleotide bias was analysed by outputting the sum of each nucleotide found in the FASTA files. Output files were processed using R Studio 1.1.463 (262) running on R 3.5.3 (263) to produce bar charts and heatmaps. The scripts are provided in Supplementary Materials Appendix AS.41 and Appendix AS.42.

2.3.15 Phylogenetic tree construction of pADAP backbones

Several core replication, partitioning, mobilization, linearization and transport genes present in all pADAP backbone variants (*repA*, *parA*, *parB*, *parC*, *mobA*, *mobB*, *mobC*, *trbA*, *trbB*, *trbC*, *traU*, *traW*, *traX*) were extracted from their respective strains. A global alignment was made using Clustal Omega using default settings (258). PhyML (259) was used, using theHKY85 nucleotide substitution model (260), to generate 100 bootstrapped trees for each alignment, a maximum likelihood tree was generated from each set respectively. The ML trees were extracted and the SumTree program v4.0.0 (269), part of the DendroPy v4.0.0 python library (270), was used to summarize all ML trees into one single consensus tree. Tree branches and labels were coloured using CorelDraw X8 v19.0.0.0328 (261) graphical software. Full alignments, alignment tree files and the final tree generated by SumTree can be viewed in Supplementary Materials S.43.

2.3.16 Data avialability

Sequencing data generated and analysed in this study have been submitted to NCBI under GenBank accession no. MT039141-MT039228. Scripts, where applicable, were deposited on https://github.com/IamIamI/pADAP_project/ . Additional data was submitted to the Lincoln University data repository at https://www.doi.org/10.25400/lincolnuninz.12736217

2.4 Other methods

2.4.1 Bioassay assessments

Pathogenicity was determined based on maximum dose oral challenge bioassays as described by Glare et al. (81). Bioassays were performed on 2nd and 3rd instar larvae of several New Zealand endemic beetle species including grass grub (*Costelytra giveni*), red headed cockchafer (*Adoryphorus couloni*), manuka beetle (*Pyronata festiva*) and Tasmanian grass grub (*Acrossidius tasmaniae*). Example pictures of the larvae in comparison to grass grub can be found in Supplementary Materials S.9. Healthy grubs were fed ~3 mm^3 carrot dices, coated in o/n culture of bioassay isolates. Grubs were checked on day 3, 6 and 12 to see if there was phenotypical change (i.e. amber discoloration because of Sep induced gut clearing). At day 3 and 6, after determining discoloration and feeding percentage, grubs were transferred to a clean tray and were provided an uninoculated carrot cube, to observe is there is was a cessation of feeding (induced by Afp). Final observation made on day 12 were used to determine bioactivity of the isolate. An example of a bioassay layout can be found in Supplementary Materials S.9. Isolates were classified as followed : hyper-pathogenic (causing ≥75% mortality in 12 days); chronic (≥75% amber disease) and mixed (<75% mortality, <75% amber disease, ≥75% affected) (271). Isolates showing <75% disease response do not satisfy Koch's third postulate that states that bacteria should cause the described disease (Amber disease or Hyper pathogenicity) upon introduction to a healthy organism, and are therefore direct correlation of the bacteria to the disease cannot be established and thus are called non-pathogenic.

2.4.2 Statistical analysis of bioassay data

Bioassay data were summarized by counting the total amount of diseased, dead and healthy grubs at day 12 for each bioassay as described by (271). The counts for each bioassay per strain were summed to generate total mortality and disease counts. Standard error percentage was calculated using the following calculation: $\sqrt{\left(\frac{\frac{X}{100}*\left(1-\frac{X}{100}\right)}{(n-1)}\right)}*100\%$

Strains are deemed pathogenic if less than 25% of the control (i.e. grubs fed with uncontaminated carrot cubes) are diseased, and 75% or more of the tested grubs (i.e. grubs fed with inoculated carrot cubes) display disease as described by Jackson et al. (271). As mortality rate was not documented in earlier publications, the “Hyper” pathotype was added for strains displaying 75% or more dead grubs at day 12 and the “Mixed” pathotype for strains that did not have >=75% death or disease, but still had 75% or more grubs being affected overall.

2.4.3 Maceration of grass grub for extraction of bacteria

To determine rates of plasmid loss and HGT of pathogenic regions of the pADAP plasmid inside of the grass grub host, several bioassayed larvae were selected for extraction and put into a 50 ml Tarsons® SPINWIN conical bottom centrifuge. Grubs were washed with ~10 ml Ethanol 70% to remove external contaminants and dirt. Cleaned grubs were removed from the tube and dried on paper towels. Grubs were washed a second time in a 50 ml tube with sterile UPW to remove any alcohol residue. After washing the grubs were transferred to a clean 1.5 ml microcentrifuge tubes.

Depending on the size of the grub, ~500 µl of sterile UPW was added. Using a sterile plastic pestle, grubs were macerated in the liquid until the liquid turned a consistent colour and texture. Approximately 5 µl from the macerated liquid was transferred to a 28 ml McCartney bottles containing 3 ml of LB-broth and briefly vortexed. The resuspended liquid was diluted 1:10 by transferring 5 µl into a clean microcentrifuge tube and adding 45 µl of sterile LB. The liquid was independently plated out in 1:1 and 1:10 dilutions onto separate LB-agar plates and incubated at 30°C.

Chapter 3 Plasmidome assembly

3.1 Sequenced samples

In 2011 Hurst et al. (83) published the first complete nucleotide sequence of the pADAP plasmid (81) obtained from the A1MO2 strain (77), formerly referred to as A1 (18). Since 2011 several parts of the virulence determinants of pADAP and the hyper-pathogenic AGR96X strain (69) underwent re-sequencing, but never the entire plasmid due to the price of sequencing in the past. With the severe drop in costs of sequencing in recent years, this study was set up to determine genetic diversity of pADAP plasmids. The main focus was on determining whether Pathogenicity Associated Islands (PAIs) were actively transferable and if there were common Horizontal Gene Transfer (HGT) drivers associated with acquisition or loss of these PAIs. In this study the genomic DNA (gDNA) of 76 bacterial isolates was sequenced, as outlined in Section 2.3.1, some with atypical virulence response in grass grub or with interesting phenotypical characteristics previously described by Dodd (94). These characteristics included bacterial species, grub mortality and disease rates, number of plasmids, estimated plasmid sizes and geographical location. An attempt was made to sequence an equal amount of virulent strains to non-virulent strains per species, although this was not possible for all *Serratia* species. The majority of sequenced isolates were *S. proteamaculans* as these had more phenotypical diversity as well as a higher genetic diversity based on previous Restriction fragment length polymorphism (RFLP) examinations (94). Apart from *Serratia* strains, one *Yersinia* strain noted by Dodd et al. (90) to have the Sep virulence determinant on a non-pADAP plasmid, was sequenced as well to determine if it shares HGT elements that facilitated the transfer from pADAP to this novel plasmid.

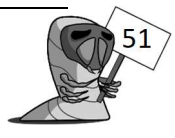
A summary of the types of isolates chosen can be found in Table 2-1. The complete overview including strain numbers, locational data, summarized bioassay data results, can be found in Supplementary Materials S.1.

3.2 Pathogenicity of isolates

Pathogenicity was determined based on maximum dose oral challenge bioassay (81) previously undertaken by Dodd et al. in 2003 (94). Bioassays of these isolates were repeated in this study, as described in Section 2.4.1, or performed by Hurst [unpublished data], to confirm accurate description of phenotype, as in 2003 it was assumed all larval disease was associated with the chronic Amber disease as hyper-pathogenicity had not yet been observed. Classifiers were determined based on disease, mortality and overall percentage of affected larvae at day 12 of bioassays as outlined 2.4.2. Results of bioassays can be observed in Table 3-1.

Table 3-1: Bioassay data for all isolates.[1] Isolates bioassayed by M. Hurst or A. Beatty [unpublished data]. Values are averaged over triplicate assays plus/minus their standard error.

Strain	Disease%	Mortality%	Affected%	Classification
A1M02	83.3 ± 4.1	3.6 ± 2	86.9 ± 3.7	Chronic
1[1]	25 ± 13	0 ± 0	25 ± 13	Non-path
4	91.7 ± 8.3	8.3 ± 8.3	100 ± 0	Chronic
142	91.7 ± 5.8	8.3 ± 5.8	100 ± 0	Chronic
143	69.4 ± 7.8	5.6 ± 3.9	75 ± 7.3	Mixed
145	95.8 ± 4.2	4.2 ± 4.2	100 ± 0	Chronic
149	13.9 ± 5.8	11.1 ± 5.3	25 ± 7.3	Non-path
158	33 ± 10	0 ± 0	33 ± 10	Non-path
163	16.7 ± 7.8	33.3 ± 9.8	50 ± 10.4	Non-path
176	95.8 ± 4.2	4.2 ± 4.2	100 ± 0	Chronic
210	90.9 ± 6.3	4.5 ± 4.5	95.5 ± 4.5	Chronic
299	50 ± 8.5	5.6 ± 3.9	55.6 ± 8.4	Non-path
336	33.3 ± 9.8	0 ± 0	33.3 ± 9.8	Non-path
345	58.3 ± 10.3	4.2 ± 4.2	62.5 ± 10.1	Non-path
376	26.1 ± 9.4	17.4 ± 8.1	43.5 ± 10.6	Non-path
377	54.2 ± 10.4	12.5 ± 6.9	66.7 ± 9.8	Non-path
398	91.7 ± 5.8	8.3 ± 5.8	100 ± 0	Chronic
465	100 ± 0	0 ± 0	100 ± 0	Chronic
626	83.3 ± 7.8	4.2 ± 4.2	87.5 ± 6.9	Chronic
1048	68.8 ± 6.8	14.6 ± 5.1	83.3 ± 5.4	Mixed
1071	41.7 ± 10.3	4.2 ± 4.2	45.8 ± 10.4	Non-path
1100	87.5 ± 6.9	8.3 ± 5.8	95.8 ± 4.2	Chronic
1129	20.8 ± 8.5	75 ± 9	95.8 ± 4.2	Hyper
1137	0 ± 0	2.8 ± 2.8	2.8 ± 2.8	Non-path
1457	0 ± 0	5.6 ± 3.9	5.6 ± 3.9	Non-path
1769	22.2 ± 7	2.8 ± 2.8	25 ± 7.3	Non-path
1770	54.2 ± 10.4	4.2 ± 4.2	58.3 ± 10.3	Non-path
1772	45.8 ± 10.4	16.7 ± 7.8	62.5 ± 10.1	Non-path
25E[1]	17.4 ± 5.7	50 ± 7.5	67.4 ± 7	Non-path
G[1]	4.5 ± 4.5	68.2 ± 10.2	72.7 ± 9.7	Non-path
10novel[1]	4.5 ± 4.5	81.8 ± 8.4	86.4 ± 7.5	Hyper
1a[1]	0 ± 0	80 ± 13.3	80 ± 13	Hyper
RM5[1]	80 ± 13.3	20 ± 6	100 ± 0	Chronic
LC[1]	10.7 ± 4.2	87.5 ± 4.5	98.2 ± 1.8	Hyper
Moraki_2[1]	97.2 ± 5	0 ± 0	97.2 ± 5	Chronic
AGR96X[1]	10 ± 9	90 ± 9	100 ± 0	Hyper
200931	5 ± 6.6	90 ± 9	95 ± 6.6	Hyper
Sprot5[1]	50 ± 15.1	40 ± 14.8	90 ± 9	Mixed
Diarr[1]	75 ± 13.1	20 ± 12.1	95 ± 6.6	Chronic
CfB[1]	40 ± 14.8	50 ± 15.1	90 ± 9	Mixed
SpF[1]	35 ± 8.3	60 ± 8.3	95 ± 8.3	Mixed
R10[1]	30 ± 13.8	70 ± 13.8	100 ± 0	Mixed
E[1]	37.5 ± 8.7	31.3 ± 8.3	68.8 ± 8.3	Non-path
Puna18[1]	0 ± 0	0 ± 0	0 ± 0	Non-path
12newD[1]	83.3 ± 11.2	16.7 ± 11.2	100 ± 0	Chronic
12a[1]	20 ± 12.1	50 ± 15.1	70 ± 13.8	Non-path
12d[1]	60 ± 14.8	30 ± 13.8	90 ± 9	Mixed
28F[1]	15 ± 10.8	55 ± 15	70 ± 13.8	Non-path
M[1]	80 ± 12.1	5 ± 6.6	85 ± 10.8	Chronic
K[1]	30 ± 13.8	10 ± 9	40 ± 14.8	Non-path



3.3 Geographical distribution of samples

With the goal of defining a wider distribution of genetically diverse species, isolates were also selected based on the location of the first isolation. The geographic distribution of sample can be seen in Figure 3-1.

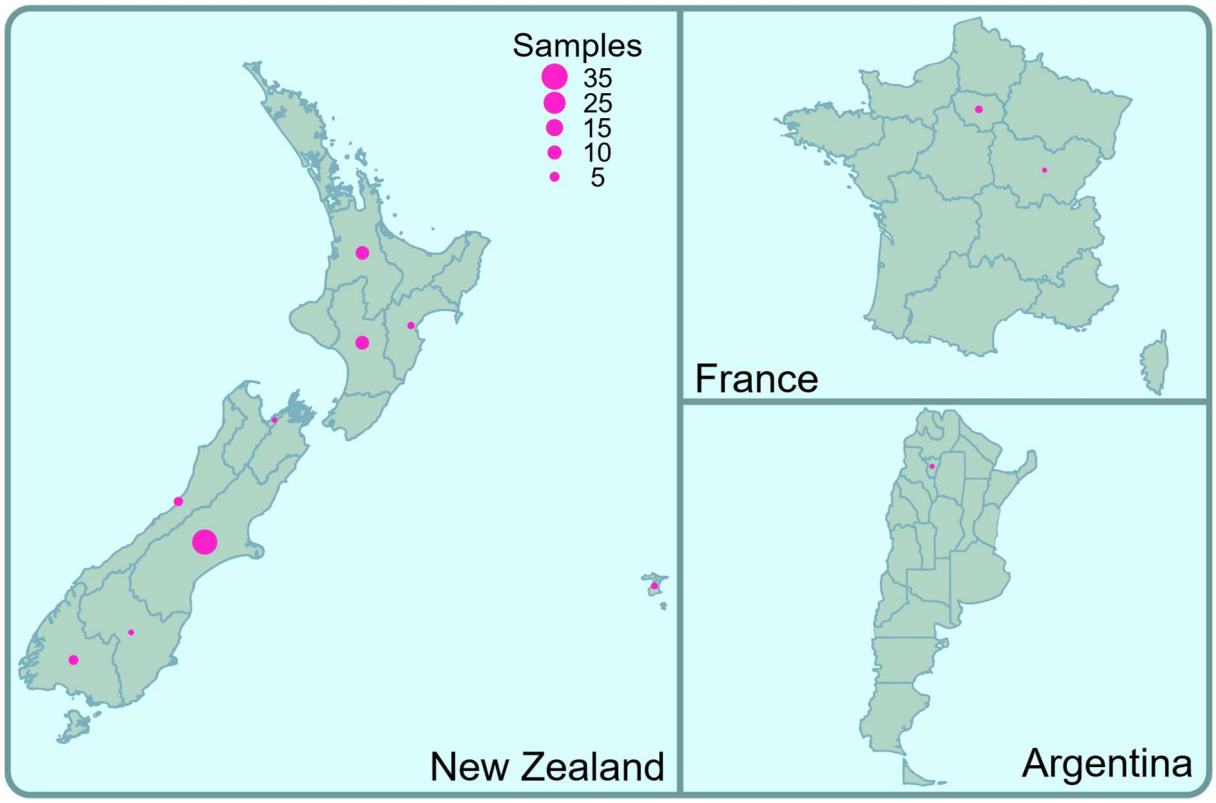


Figure 3-1; Geographical distribution of sequenced isolates. The majority of isolates were obtained from New Zealand, 2 isolates were obtained from the Pasteur institute in France and one from the Bordeaux region in France, and one isolate obtained from Tucuman region in Argentina (81). The dots are centroids for the provinces, as coordinates for some isolates prior to 1990's did not have ample description to pin point their accurate source location.

Note that *Serratia entomophila* 220, *Serratia ficaria* 457 and *Serratia grimesii* 348 isolates were sourced from France and the *Serratia proteamaculans* 495 isolate was obtained from Argentina. Although plasmid visualization gels showed megaplasmids similar to the size of pADAP in all but one of the four non-endemic strains, none were found to harbour a pADAP-type plasmid (81). In the latter part of this study, one additional dataset was added for a *Serratia marcescens* isolate designated WVU-005, sequenced by Ranjbaran et al. in July 2019 [PRJNA545504]. The WVU-005 isolate was derived from a clinical sample from the West Virginia University in Morgantown, USA.

3.4 **Plasmid contig extraction and annotation**

Of the 76 isolates that were sequenced, as outlined in Section 2.3.1, and assembled as described in Section 2.3.3 and Section 2.3.4, 27 were found to contain multiple plasmids, some even containing multiple megaplasmids. Six strains that were sequenced were assumed to not have any plasmids based on megaplasmid visualization gels, and analysis of the sequencing data confirmed this. Four of the isolates sequenced were of poor sequence quality, preventing proper assembly and were therefore excluded from further analysis. Obtained from the remaining 72 isolates were a total of 83 plasmids, suitable for further study, 50 of which were confirmed to be pADAP or pADAP-like, defined by the plasmid backbone. Further explanation on the process of extraction of plasmid contigs can be read in Section 2.3.5. Several plasmid contigs were PCR validated as detailed in Section 2.3.6 to verify accuracy of the assembly pipeline. A breakdown of the plasmid carriage per species can be seen in Table 3-2.

Table 3-2: An overview of the plasmid content of isolates, per species.

Genus	Species	Isolates sequenced	Rejected due to quality	Plasmid bearing	Multi-plasmid bearing	Total Plasmids	pADAP-backbone plasmids
Serratia	*entomophila*	16	1	13	4	19	10
	ficaria	1	-	-	-	-	-
	grimesii	1	-	1	0	1	0
	liquefaciens	3	-	2	0	2	2
	marcescens	3	-	3	1	3	0
	proteamaculans	51	3	48	10	55	39
Yersinia	*frederiksenii*	1	-	1	1	4	0
Total		76	4	68	16	84	51

Plasmids contigs that were obtained from the assemblies were annotated by Prokka (249) (full method is outlined in Section 2.3.8) to generate CDS features (a.k.a. ORF or genes), using the bacterial RefSeq protein database (272) as a reference database. Additionally, three plasmids were added from NCBI, the pADAP reference plasmid [NC002523] (83) and two plasmids belonging to the recently sequenced *S. marcescens* strain WVU-005, sequenced by Ranjbaran et al. [PRJNA545504]. The WVU-005-1 plasmid shares portions of the pADAP backbone and is therefore included in further analysis to provide evolutionary context to the pADAP family of plasmids. The full overview of all strains sequenced including number of plasmids per strain, the plasmid sizes predicted by Dodd (94) and Hurst [unpublished data], plasmid sizes obtained from the assemblies, the sequencing technology, locations of strain isolation to the highest accuracy that was available, gene counts per plasmid, and other characteristics, can be found in Supplementary Materials S.1.

3.5 **Determining pADAP copy number**

Before further pADAP related comparisons could be made, the plasmid copy number needed to be established to determine whether the assemblies are not affected by reads belonging to conflicting pADAP plasmids, thus affecting the proper construction of the De Bruijn graph during assembly (273). Plasmid copy number was determined by analysing the coverage data of the chromosome contigs and plasmid contigs obtained for several strains (Section 2.3.7). For this, six of the high-quality PacBio assemblies were taken and, their respective Illumina reads were mapped to them and a coverage plot was produced (Figure 3-2).

The coverage plot in Figure 3-2, clearly shows that the average coverage per nucleotide on the plasmid contig(s) is equal to the coverage of nucleotides on the chromosome contig, which indicates that plasmids were sequenced in the same depth as the chromosome.

These results suggest there was a similar ratio of plasmid and chromosomal template presence during sequencing, i.e. one plasmid template per one chromosome template. In case of a multicopy vector you would expect to see the complete distribution shift to twice the coverage or more of that of the chromosome, depending on the number of plasmid copies. As there would be at least twice as much template available in the sample to sequence, in case of multi-copy vectors, it would yield twice as much reads. Thus, multi-copy vectors would have an increased average coverage, which was not observed here. Several *in vitro* attempts were made to conjugate antibiotic tagged pADAP variants, into strains already carrying another tagged pADAP variant. These experiments, however, all proved unsuccessful and corroborates self-incompatibly of the pADAP plasmid. One mechanism that is most likely causal of this self-incompatibility of pADAP plasmids is handcuffing as outlined in Section 4.1.

After having established that pADAP is a single copy megaplasmid, it was reasonable to assume that plasmid contigs do not contain any assembly artefacts caused by diverging de Bruijn graph paths that could have arisen from two or more diverging pADAP type plasmids being present in one single dataset. With this fact established, it was now possible to do some comparative plasmidomics. The most conserved part of plasmids is their replication region; therefore, this was the first region that was studied.

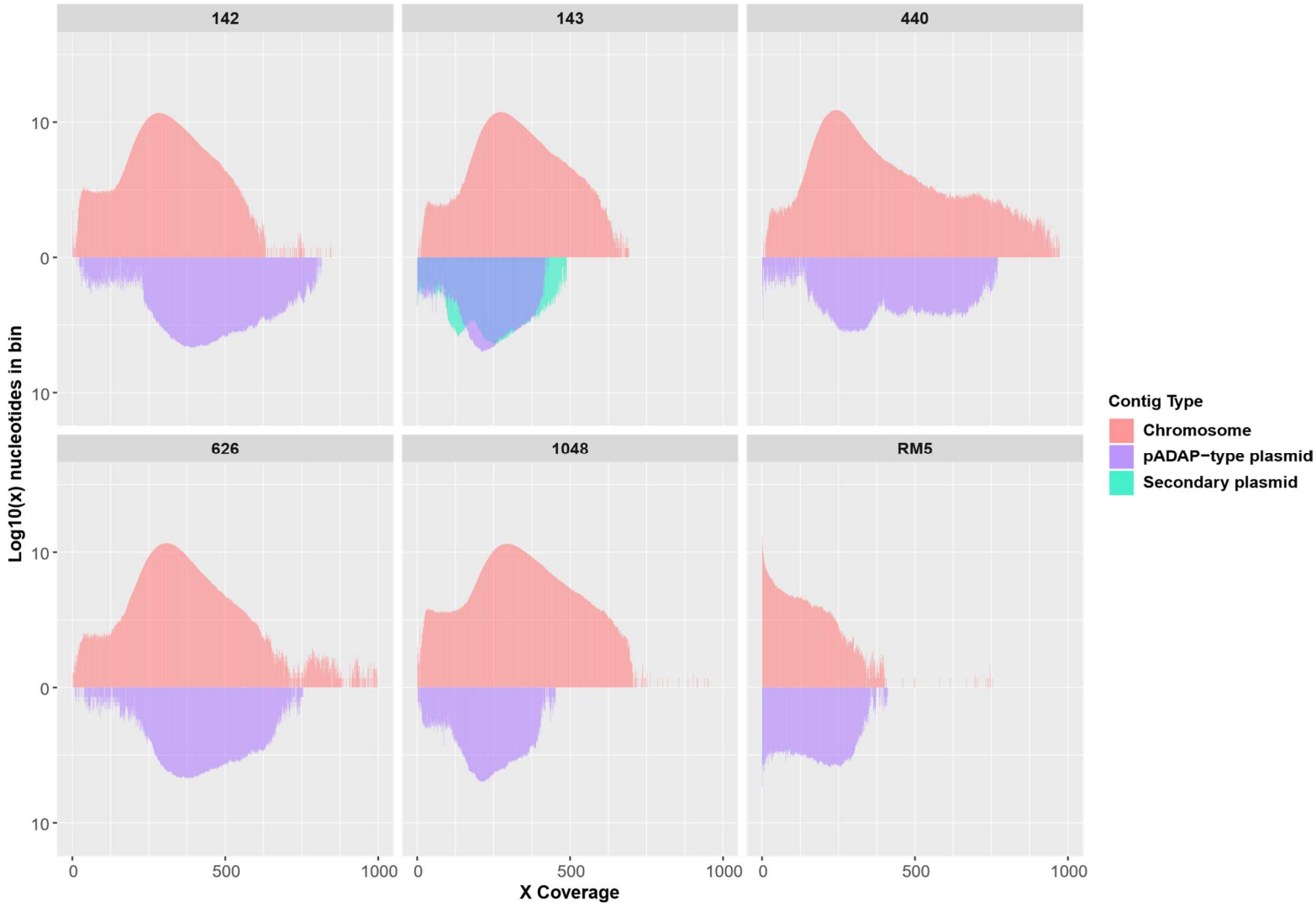


Figure 3-2: Coverage plot of the six samples sequenced using both the PacBio and the Illumina platform. Illumina reads were mapped against the high quality PacBio contigs to determine the coverage. These coverages were then plotted in bins, with the chromosomes mapped in an upward direction and the plasmids in a downward direction. The number of nucleotides on a log scale (y-axis) per coverage (x-axis).

Chapter 4 Comparison of pADAP backbone regions

4.1 RepA orthology analysis

The RepA protein found in the conserved backbone of pADAP plasmids is required for self-replication of the plasmid. In normal circumstances, a RepA protein can upregulate its own transcription by binding to a series of DNA repeats called iterons, in its own promotor region (Section 4.2). The DnaA replication initiation factor, normally associated with the ori of the host cells chromosome, is able to interact with the iteron bound RepA proteins, and bind to the DnaA-box upstream of the *repA* gene (274, 275). The RepA protein has also been shown to interact with other initiation factors such as DnaB and DnaC (276). These components of the primosome are required to initiate replication.

Several RepA orthologs were obtained from the RefSeq database (250) in order to determine the evolutionary origin of the pADAP RepA based plasmids. Top twenty RepA ortholog hits from diverse species were extracted, aligned and used to construct a phylogenetic tree. The resultant phylogenetic tree shows the relationship of replication genes in other bacteria, and it shows the unique position pADAP has within this family of RepA-based plasmids. The tree can be viewed in circular format in Figure 4-1 and the corresponding amino acid sequence similarity for RepA can be seen in Figure 4-2. Detailed description of the method used to generate the tree and distance matrix can be found in Section 2.3.9 and Section 2.3.10 respectively.

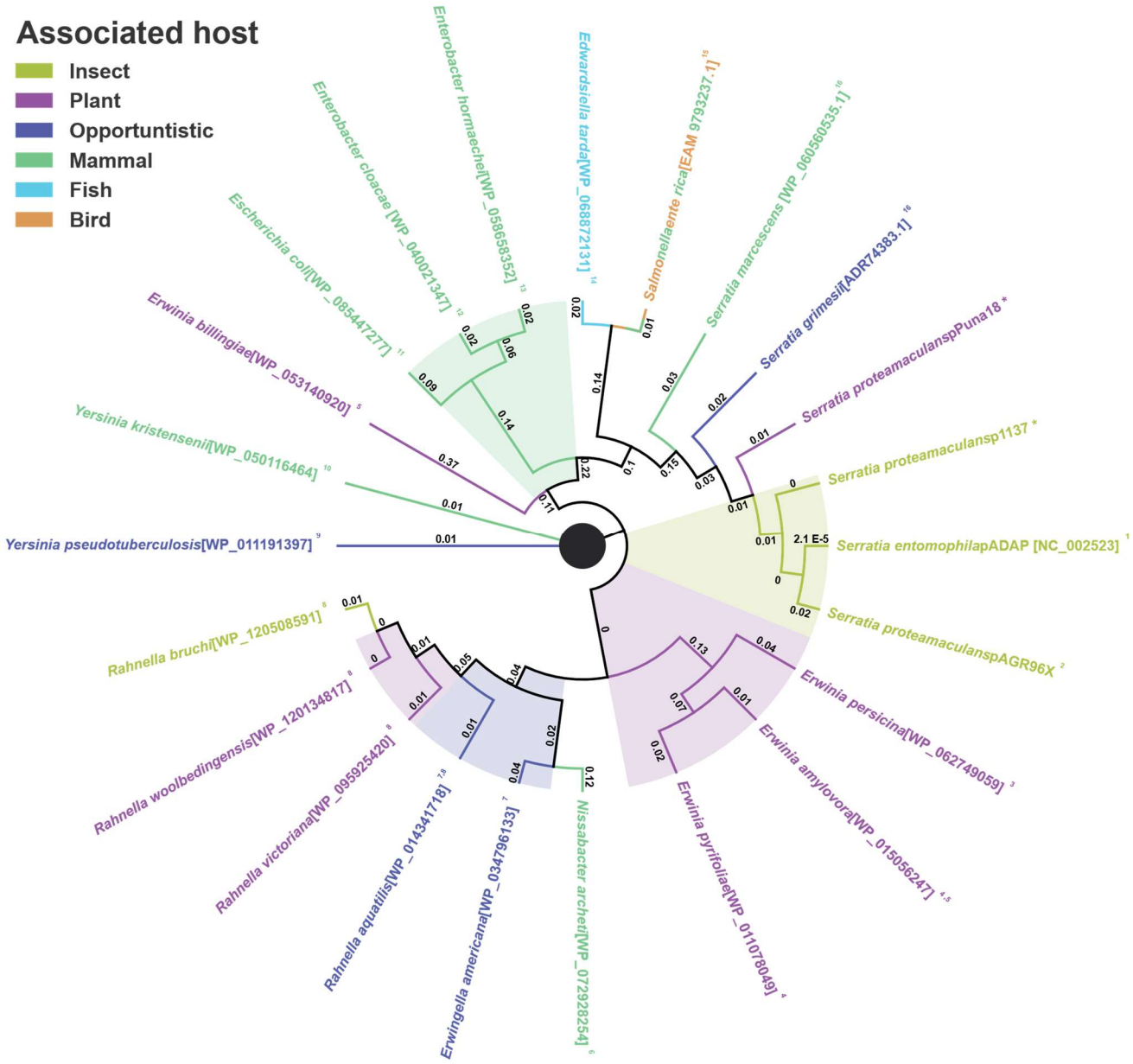


Figure 4-1: A phylogenetic tree of RepA pADAP as well as its 20 closest orthologues. The pADAP RepA is the only RepA ortholog associated with entomopathogenic bacteria. *Rahnella bruchi* is found in insect guts, but has not been described as being pathogenic. References used to determine closest host range of bacterium; * This study, [1] Grimont et al. (18), [2] Hurst et al. (69), [3] Zhang et al. (277), [4] Rhim et al. (278), [5] Kube et al. (279), [6] Mlaga et al. (280), [7] Brenner et al. (281), [8] Brady et al. (282), [9] Martínez-Chavarría et al. (283), [10] Robins-Browne et al. (284), [11] Souza et al. (285), [12] Lui et al. (286), [13] Townsend et al. (287), [14] Wang et al (288), [15] Rabsch et al. (289), [16] Mahlen et al. (290).

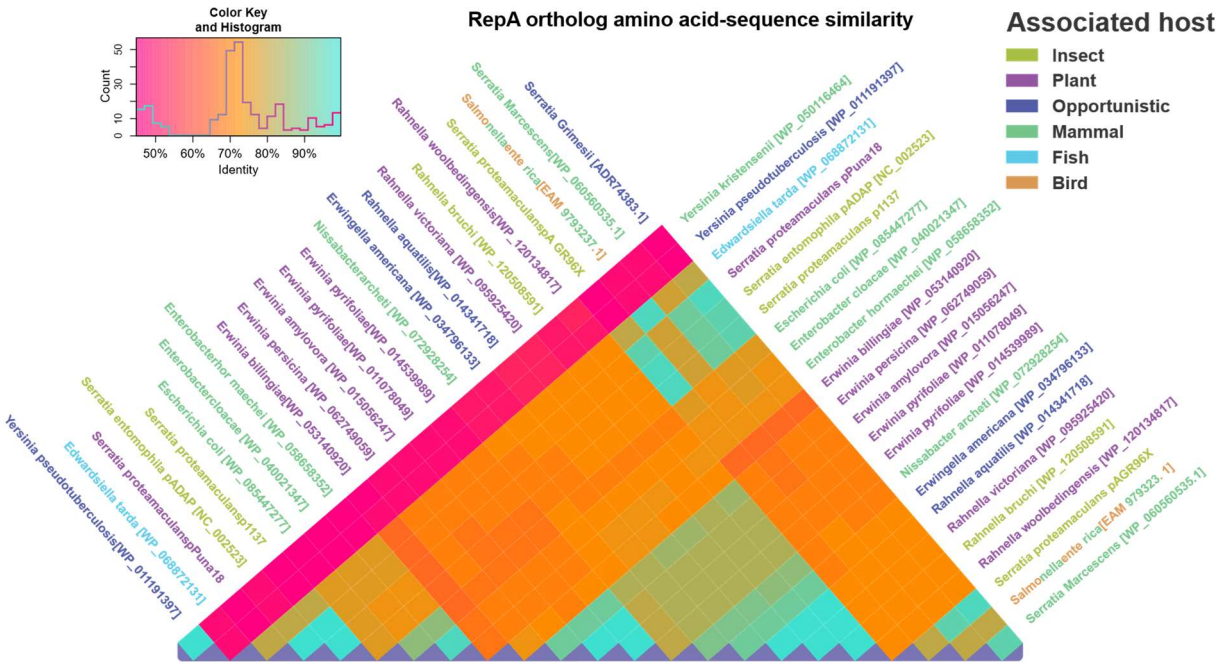


Figure 4-2: Heat map of a distance matrix of the *repA* ortholog AA sequences. This figure represents an all against all AA sequence similarity matrix.

A literature study was undertaken on each bacterial species associated with the respective RepA ortholog, to determine their closest associated host species (Figure 4-1). From this analysis, the pADAP RepA appears to be the only protein in this dataset that is associated with insecticidal bacteria (Figure 4-2). This suggests that within this group of closely relate plasmids based on their replication gene, the pADAP variant carrying isolates are the only ones who appear to have co-evolved with an insect host. Full descriptive summary of reference sequences used, and literature surveyed, as well as full region alignments can be found in Supplementary Materials S.13. It must be said however, that the absence of literature associating these bacterial species with insecticidal activity does not mean that their respective plasmids or associated chromosomes do not carry insecticidal elements.

4.2 **RepA associated iterons**

As outlined in Section 3.5, based on sequence coverage, pADAP is predicted to be a single copy plasmid. The single copy nature of the plasmid suggests self-incompatibility. Plasmid incompatibility is still a commonly used method to genotype plasmids (100). It is based on the principle that certain replication regions are not able to function well in the presence of plasmids with an antagonistic replication region, and so cells cannot maintain the replication of both plasmids in order to vertically transmit them to progeny during cell division (99). Based on the known Inc group primers (100), no nucleotide hits were identified in any of the pADAP plasmids, revealing that pADAP likely represents a new Inc group plasmid. However, conjugation tests with numerous known Inc group plasmids need to be performed to accurately determine the true Inc group of pADAP.

Based on the single copy nature of pADAP, the means behind self-incompatibility of the pADAP RepA is most likely handcuffing (291). Handcuffing is an event where two plasmids with similar replication genes interlock, preventing access to the replication site (292). RepA is a self-promoting protein, upon transcription it binds to iterons, which are small repeat motifs, ~12 to 15 bp in length, positioned 5' of the replication gene (293). Here, the bound RepA proteins interact with the replication initiation factors such as DnaB and DnaC (276), but most notably DnaA and initiate replication of the plasmid (275). If all iterons are RepA bound and a second plasmid is present that is in the same state, it is possible for the iteron bound RepA monomers, to interact with iteron bound RepA monomers from the other plasmid copy and dimerize, essentially "handcuffing" the replication region of the two plasmids together (274). Once "handcuffed", further interaction with the upstream region of *repA* is blocked and further replication of the plasmids is stopped (291-293). This handcuffing prevents what is called as the 'dimer catastrophe', which is a situation where clonal multi-copy plasmids that are in a dimer state after replication of during recombination, keep replicating. Because the monomers have only one origin of replication, the dimerized plasmids are more likely to replicate due to them now encoding two or more origins of replication. This causes a cascade of plasmid dimers and results in a metabolic overload to the cell, as well as halting proper plasmid segregation towards progeny during cell division (294). By preventing replication through handcuffing, dimerization of the plasmid cannot occur. If the handcuffed plasmids are already in a dimerized state, they can be resolved through site-specific recombination. The pADAP plasmids contain a XerCD domain containing ORF called Int2 (Section 4.4), short for integrase 2 that ensures that dimerized plasmids can be recombined at the XerCD site to form monomers (295, 296). The handcuffed plasmids will remain bound until cell division causes the cytoplasm to increase substantially by which time the two plasmids are forced to separate, each into their own cell. The iteron region of the pADAP RepA was previously described by Hurst et al. (83) and was confirmed in this study to be present in all pADAP plasmids and their positions conserved.

As RepA is highly conserved in several genera, the iteron analysis was extended to encompass several of the upstream regions of the previously mentioned RepA orthologs. Not all RepA orthologs had an associated plasmid present in the RefSeq nr database, and therefore the set of plasmids analysed was smaller than the set of RepA proteins used to construct the phylogenetic tree. The results of the structural conservation of the iteron motifs upstream of RepA orthologs can be seen in Figure 4-3.

The plasmids with a homologous *repA* gene most likely have a similar promotor region that contains iterons with a shared motif (Figure 4-3). This led to the hypothesis that pADAP does not handcuff in the presence of close pADAP variants but also more distant RepA encoding plasmids. The inability to carry more than two of these RepA encoding plasmids at a time make them incompatible in a population as during segregation each progeny will only get one of the two plasmids. This makes pADAP not only self-incompatible but also incompatible with most other *repA* plasmids. In the environment, this self-incompatibility would severely limit the opportunity for co-existence with another RepA encoding plasmid and therefore limits the opportunity for recombination with a similar pADAP-type. In an evolutionary context, this may have restricted these RepA plasmids to stop recombining, forcing them towards specialization through natural selection driven means. It also shows a good link between co-evolution of the plasmid with the bacterial host.

Alignments and description of iteron regions can be found in Supplementary Materials S.14 respectively.

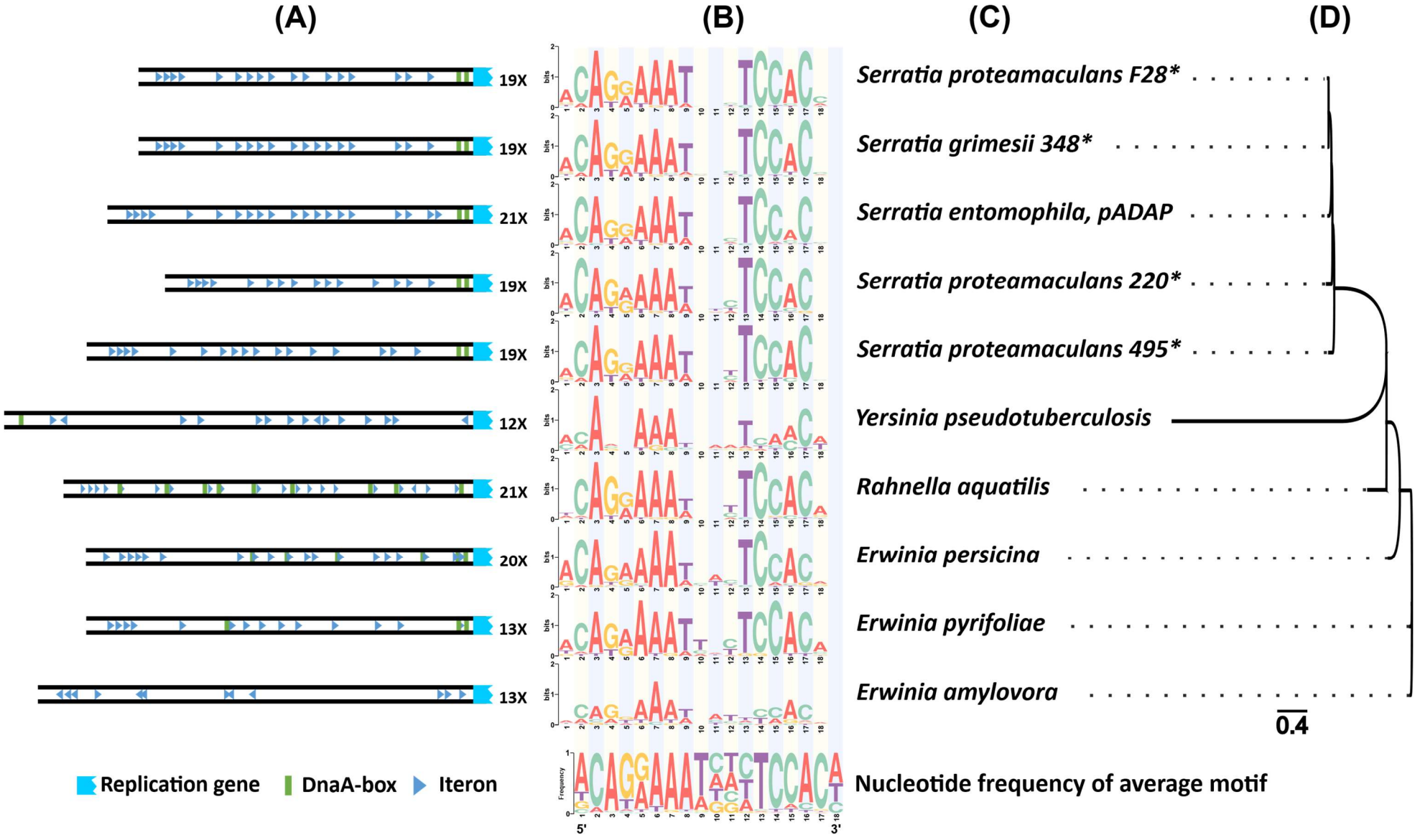

(A)
(B)
(C)
(D)
19X
19X
21X
19X
19X
12X
21X
20X
13X
13X
Replication gene
DnaA-box
Iteron
Serratia proteamaculans F28*
Serratia grimesii 348*
Serratia entomophila, pADAP
Serratia proteamaculans 220*
Serratia proteamaculans 495*
Yersinia pseudotuberculosis
Rahnella aquatilis
Erwinia persicina
Erwinia pyrifoliae
Erwinia amylovora
0.4
Nucleotide frequency of average motif
5'
3'

Figure 4-3: A schematic overview of a conserved iteron motif found upstream of the *repA* replication gene. This iteron motif was found on pADAP by Hurst et al. (83). Motifs were found to be highly conserved in all pADAP plasmids, both in the motif and in their orientation upstream of the replication gene. Iterons are associated with both promoting expression of the *repA* replication gene through autoregulation, but in presence of two or more plasmids are also known to be involved in "handcuffing". A) Pattern of iterons and DnaA-boxes found upstream of the replication genes in several plasmids that share sequence homology with RepA found on pADAP. B) Sequence logo of the conserved iterons, the triple A and the TCC nucleotides are the highest conserved within the motifs, sequence logos generated using WebLogo (297). C) The names of the species that contained the compared plasmids, full plasmid name and description can be found in Supplementary Materials S.14. D) the evolutionary distance between each RepA replication protein based on PhyML generated ML phylogenetic tree (259).
*** Sequences of non-pADAP plasmids found in this study, that contain a replication gene with homologous iteron region but do not contain the pADAP backbone, or portions of it.**

4.3 pADAP backbone

A key aim of this project was to determine whether previously sequenced virulence determinants published by Hurst et al. (69, 83) were associated with any transposable mechanism that could actively acquire or remove these PAIs. To discern the reason behind the dissimilarities of the virulence regions, first, the capturing apparatus, i.e. the plasmid backbone, needed to be analysed. The backbone consists of the replication and conjugation region as previously described by Hurst et al (83), and was used to determine whether a plasmid is a pADAP-type (Figure 1-5).

From the 72 successfully sequenced strains 52 circular pADAP plasmid contigs were obtained. The first step in comparing the plasmid sequences was to establish the most conserved "start" of the plasmid. A full manually curated list of deviating ORFs of all pADAP-type plasmid backbones is provided in Supplementary Materials S.40.

4.4 The conserved "start" of the pADAP-type plasmids

Int2 was believed to be a potential integrase upstream of the *repA* gene (83), however, updated ortholog hits in the RefSeq database revealed it to be a recombinase (further discussed in Section 9.1). Alignment of a 5 Kb region, 5' and 3' of the *int2* / *repA* of each plasmid, revealed the element is highly conserved across all pADAP-type plasmids assessed in this study. The nucleotide alignment also showed a clear point of delimitation 524 bp upstream of the pADAP *int2* gene (Figure 4-4).

The point of delimitation was used to set the residue numbering of the plasmid, making the T in the (5' TTGTAATAAA 3') sequence the first residue number for all pADAP variants. Now that the start of all plasmids was defined to the same homologous point, it was fairly simple to visually compare the pADAP-type plasmids with one another.

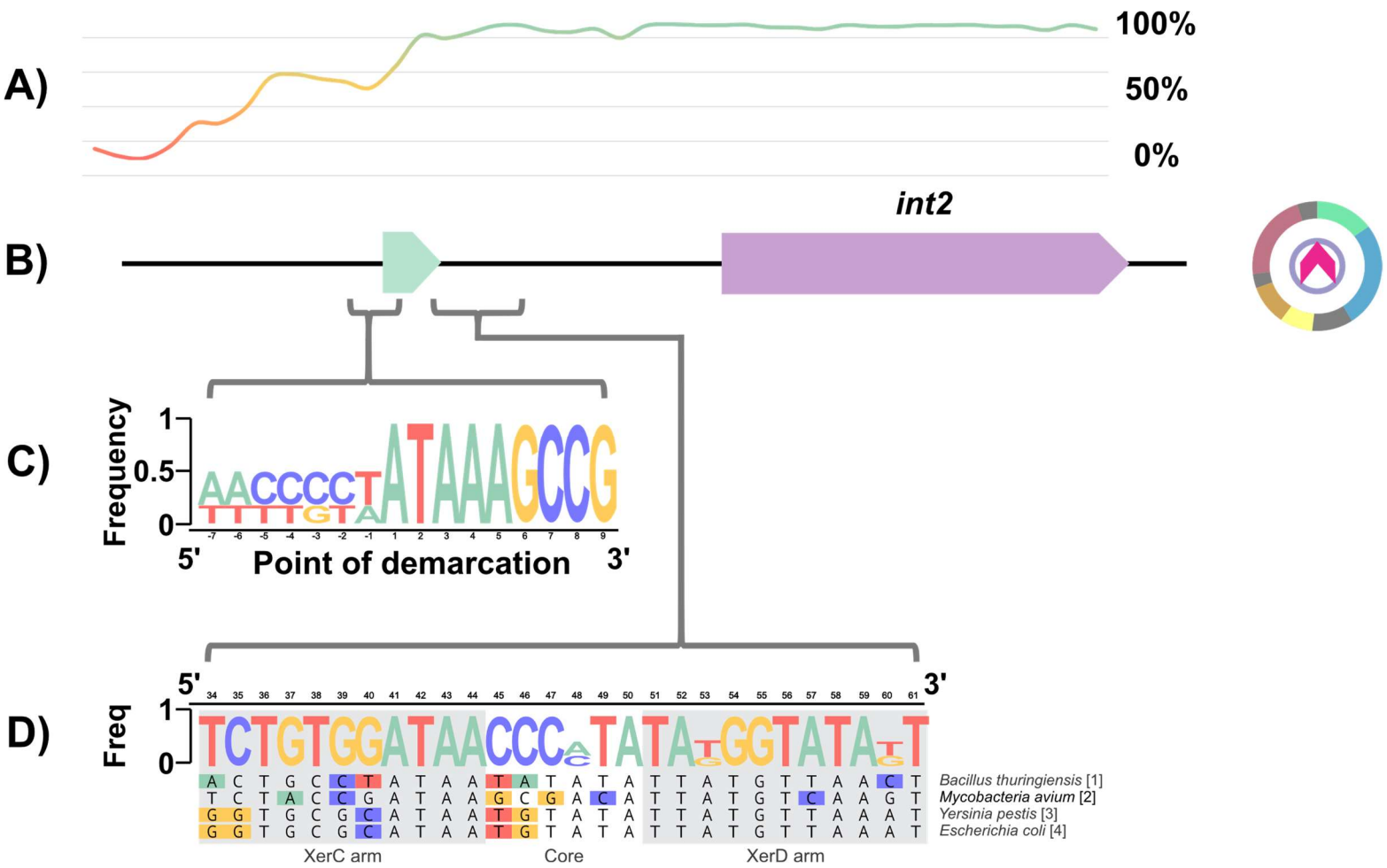


Figure 4-4: The proposed start of the pADAP type plasmids. The start of the backbone was determined to be a non-coding sequence found 524 bp upstream of the tyrosine recombinase encoding *int2* gene in pADAP [NC002523]. The region between this point of delimitation and the *int2* gene, does not contain any repeats. A) percentage of nucleotide identity between all pADAP-type plasmids analysed in this project. B) Orientation between the point of delimitation and *int2*. C) Frequency sequence logo plot showing the conservation of nucleotides for this delimitation point, an overall low frequency means that gaps were introduced into the alignment. D) The dif recombination site needed for site specific recombination. Sequence logos generated using WebLogo (297).

4.5 pADAP BRIGs analysis

The first attempt at plasmid comparisons was undertaken through the use of the Blast Ring Image Generator (BRIG) tool (264) (full method is detailed in Section 2.3.10). BRIG takes query sequences and performs a Blast+ (245) alignment against a reference sequence. The reference used in the BRIG analysis was the pADAP reference plasmid [NC_002523]. The tool then outputs a circular figure that shows which areas of the query sequences are similar to the reference sequence. It is important to keep in mind that this does not show what the areas of dissimilarity in the queries contain (indels, genomic reshuffling, HGT acquired genetic material, etc), making the figure only half the story (Figure 4-5).

.

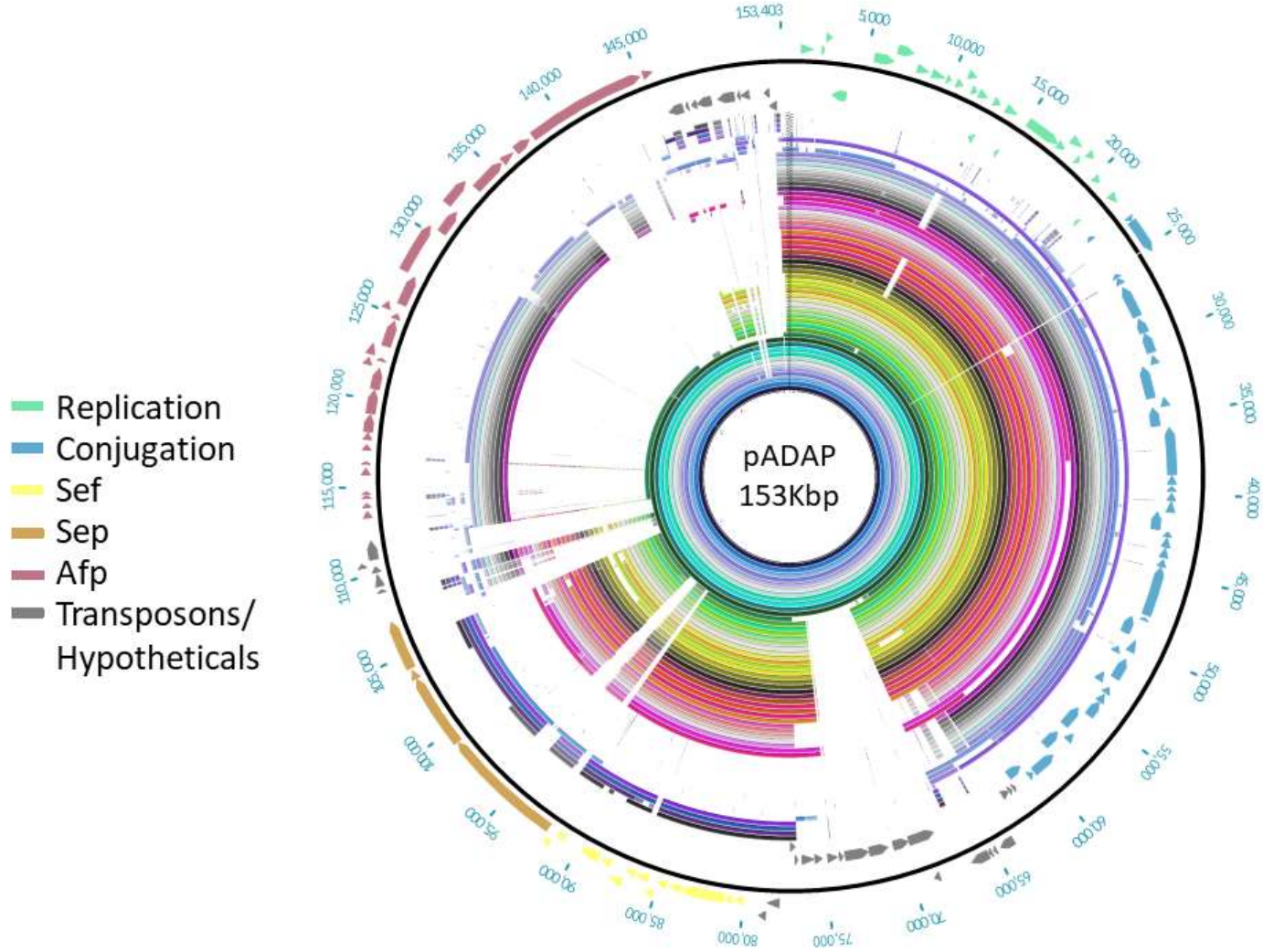


Figure 4-5: A Blast Ring Image Generator image of all the 57 pADAP type plasmids. The reference used was the A1MO2 strain pADAP [NC002523] sequence obtained from NCBI.

Based on the annotations of the plasmid encoded ORFs and the information of the BRIGs analysis 52 plasmids, out of the 84 total plasmids, were deemed pADAP or pADAP-like as they contained most of the pADAP backbone, spanning from the point of delimitation to the *pilL* ORF. Of the remaining 34 non-pADAP plasmids, 5 were found to have some pADAP elements, most of which were the Sep region or fragments thereof. A breakdown of the regions found based on the BRIGs analysis and manual inspection can be seen in Figure 4-6.

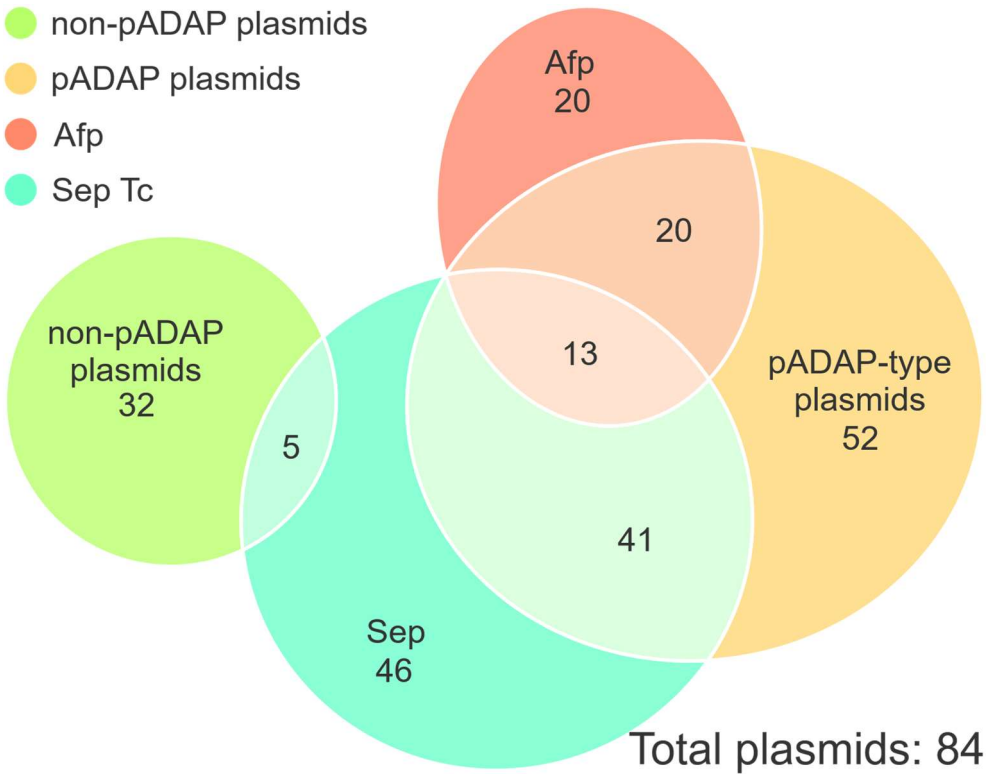


Figure 4-6: A Venn diagram showing the summary of plasmids obtained from sequenced isolates. The diagram shows the number of found plasmids, and if they contained pADAP markers such as the pADAP backbone, the Afp particle associated with antifeeding behaviour or the Sep region associated with gut clearance.

Although BRIG is a great tool to analyse similarities to a reference, it is not very effective at showing dissimilarities in the query sequences. Based on this limitation, the sequences were extracted, and a global alignment was made using Clustal Omega (258) in order to perform manually assessment of regions of divergence. The backbone region from the point of delimitation to the *pilL* ORF was approximately 62 Kb in size in the pADAP reference plasmid.

4.6 Backbone similarity

A DNA sequence similarity matrix was produced to show the raw value of nucleotide similarity between all pADAP variants (Section 2.3.10 contains a full description of the process). This provided an insight into the groups of backbones. Most of the *S. entomophila*-based pADAP plasmids shared a high DNA sequence identity to one another. Several outlier strains with distinct phenotypes were analysed such as pPUNA18, which was most likely a plant associated isolate as the plasmid did not appear to contain either SEP or AFP, but instead carried numerous plant associated effectors (see Section 8.4 for more details), although has a very similar backbone to the *S. entomophila* pADAP. There was also a small group of five sequenced plasmid including pG, p25E, p10novel and pE that formed a unique group (Figure 4-7).

As the DNA sequence similarity between some of the backbones was very low, it was of some interest to further define the areas of divergence. The backbone alignment was manually assessed, and several sites of divergence were found that either had lost or acquired novel ORFs. Most of these regions, as outlined in the subsequent sections, did not seem to code for anything that could be considered to substantially alter the functionality of the plasmid or the bacterial host. These regions however, do offer an insight into how HGT affects plasmids and how natural selection further divided the pADAP plasmid family.

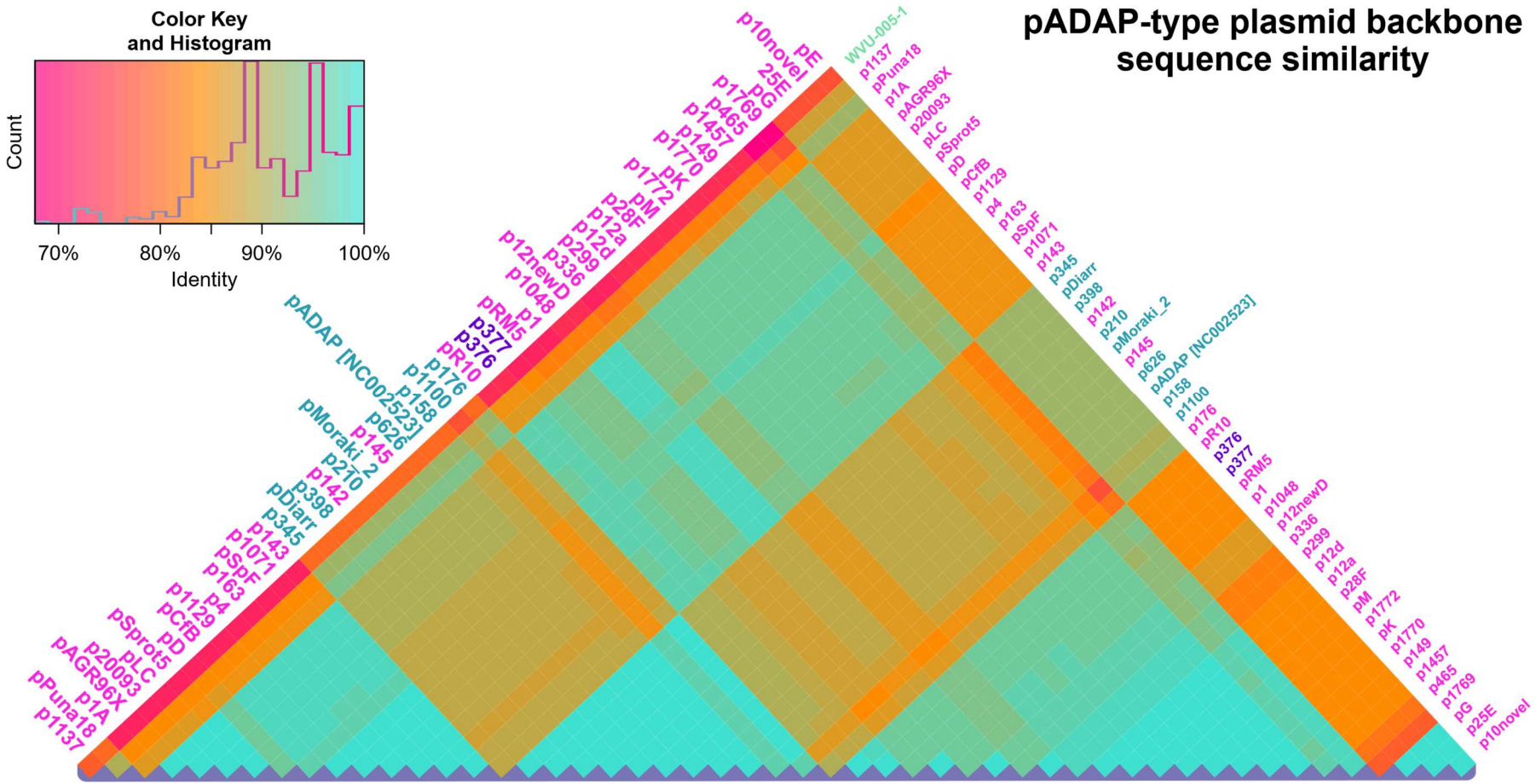


Figure 4-7: A DNA similarity matrix of the pADAP backbones of all 52 pADAP-like plasmids. The backbone is the region between the conserved point of delimitation (start) of the plasmid (Figure 4-4) and the *pilL* gene. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens* and pink *S. proteamaculans*.

4.7 The *sea36-areA* region

The first divergent region identified spanned ~2.5 Kb and comprised three *sea* ORFs and an *areA* gene. ORFs were labelled with *sea* in the 2011 pADAP annotation to denote unique hypothetical ORFs found on pADAP, and it stands for *Serratia entomophila* annotation (sea). For this study the sea ORFs were re-annotated, their function described in the text and a detailed summary can be found in Supplementary Materials S.8. The *sea36-areA* region is depicted in Figure 4-8 and the corresponding isolate per *sea36-areA* type in (Table 4-1).

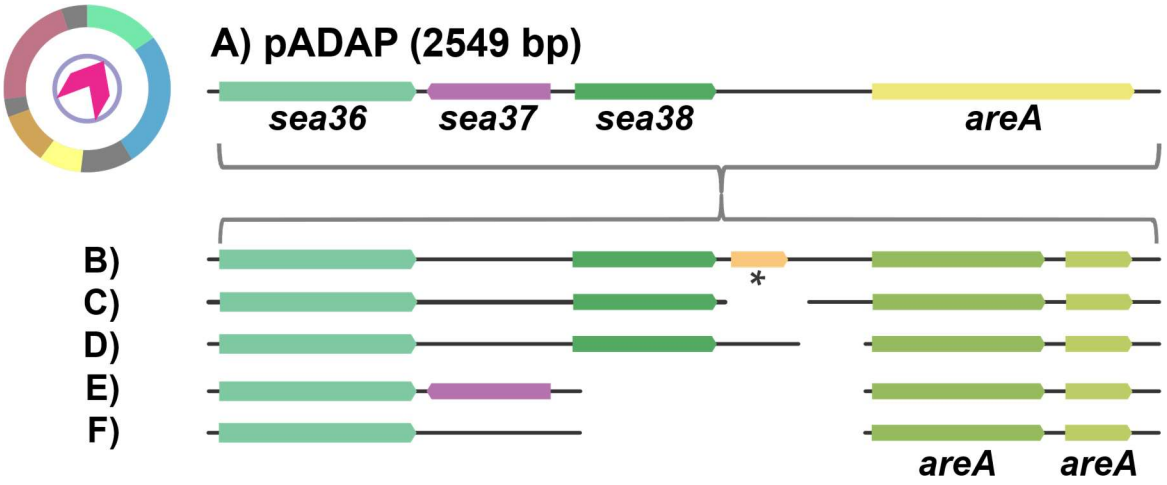


Figure 4-8: The *sea36-areA* nucleotide region. This region is approximately 2.5 Kb in size and comprises of 4 genes on the pADAP reference, *sea36* through *sea38* and the *areA* ORF. Several of the other pADAP type plasmids, however, have missing annotations due to nucleotide changes, missing regions, a cleavage of the *areA* gene and the insertion of a hypothetical gene, denoted with a *.

Table 4-1: Sea36-38, *areA* variant type per pADAP-type plasmid carrying isolate.

Type	Species		Isolate
A	*Serratia*	*entomophila*	158, 176, 398, 626, 1100, Diarr, pADAP
		proteamaculans	145
		marcescens	WVU-005-1
B		*proteamaculans*	149, 299, 336, 465, 1457, 1769, 1770, 1772, 12a, 12d, 28F, D, K, M, Puna18
C		*proteamaculans*	1, 1137, Rm5
D		*entomophila*	210, 345, Moraki_2
		proteamaculans	142, R10
		liquefaciens	376, 377
E		*proteamaculans*	143, 1048, 1071, 10novel, 12newD, 25E, CfB, E, G
F		*proteamaculans*	4, 163, 1129, 1A, 20093, SpF, LC, AGR96X, Sprot5

The *areA* gene encodes a nitrogen metabolite regulator (298) however, a BLASTP search for the AreA protein encoded by pADAP suggests that instead it is an anti-restriction ArdA protein (299), a protein that can help HGT of MGEs by preventing DNA cleavage by restriction enzymes. This gene only appears as one ORF in the pADAP plasmids found in *S. entomophila*. All other plasmids were annotated with two AreA encoding ORFs. As all plasmids assessed appear to only have one Shine-Dalgarno (SD) sequence (5' AGGAGG 3'), it's likely the split ORF is just an annotation artefact caused by an in-frame stop codon that is not in-frame in the *S. entomophila* pADAP plasmids.

Most *sea* genes are of yet to be determined function, i.e. are designated as hypothetical genes. Since the original annotation from 2011, several *sea* genes now have gene orthologues in the NCBI RefSeq database with known function. In this context, the *sea36* ORF was found to be homologous to a helix-turn-helix transcriptional regulator, and *sea38* has some AA similarity to an aerotaxis sensor receptor (Aer) (Table 4-2) (300).

The Aer is an important internal sensory protein that can react to changing energy and oxygen states in a cell (301) by signalling the cell to rapidly oxidize substrates such as ribose, galactose, maltose, malate, proline, alanine, glucose, mannitol, mannose, sorbitol, and fructose (302). The Aer protein has also been correlated to the ability of *E. coli* to colonize mouse colons, and is hypothesized to upregulate flagella related genes, keeping the cells mobile and preventing cells to remain in one place, adhere and colonize (303). Many grasses contain sugars in their roots, which is the main plant area that these S*erratia* isolates are associated with, as well as the main dietary requirement for grass grubs. Disruption in this region could have potential implications with colonization of the grass grub and thus affect the pathogenicity. Unfortunately, this was found out at the end of the study and thus no further *in vivo* assessment was undertaken.

Table 4-2: Approximation of function of ORFs in the *sea36-areA* region based on BlastX hits.

Original annotation	New annotation	Query Cover	E value	Per. Ident	Accession
sea36	MULTISPECIES: helix-turn-helix transcriptional regulator [*Serratia*]	99%	1e-111	94.59%	WP_115185061.1
sea37	MFS transporter [*Brevibacterium iodinum*]	76%	0.59	32.69%	WP_101546461.1
sea38	Aerotaxis sensor receptor, flavoprotein [*Escherichia coli*]	70%	2.00E-11	41.24%	WP_097331443.1
areA	Antirestriction protein ArdA [*Yersinia massiliensis*]	67%	3e-73	82.22%	WP_172986820.1

4.8 MGE insertions

Several other small divergent regions appear to be the result of seemingly random insertion of transposons or other MGEs with no obvious virulence associated function.

4.8.1 Insertion between sea40-42 region

Within the pADAP backbone sequence from isolate 1100, a small ~2 Kb insertion was identified in the *sea40-42* region. The *sea40* and *sea41* ORFs are two parts of the HigBA toxin anti-toxin (TA) system. A more in-depth description of the function of this and several other TAs can be found in Chapter 11. Located downstream of this TA region in isolate 1100 is a 1931 bp insertion, containing two hypothetical ORFs and a group II intron reverse transcriptase (Figure 4-9, Table 4-3). The insertion is at the 3' of the *sea41* ORF and therefore unlikely to affect the functionality of the TA operon. There appears to be no particular characteristic in prokaryotes associated with introns of this type (304). It has previously been proposed that these group II introns often act as MGEs (305). Based on this information, it is likely that this insertion has been a random acquisition with no consequence to cell or plasmid functionality

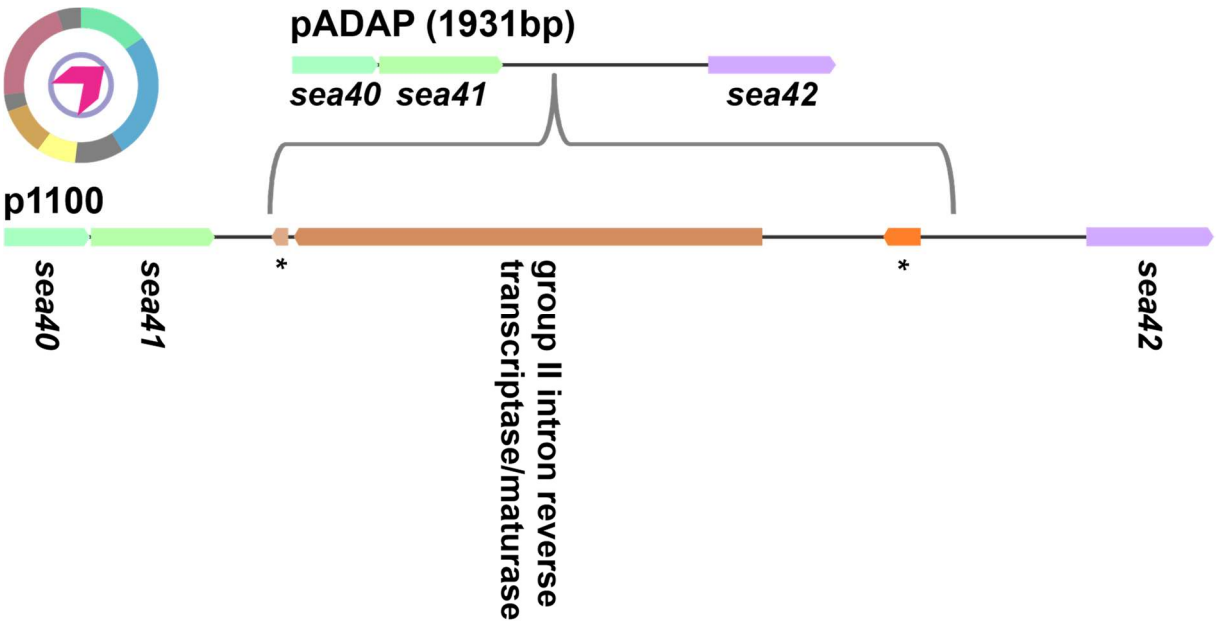


Figure 4-9: The *sea40-42* region in p1100. The pADAP sequence derived from the isolate 1100, contained a 1931 Kb insert, positioned between *sea41* and *sea42.* This region consists of a group II intron and two hypotheticals. *Sea40* is a *higB* and *sea41* is *higA.* The group II iteron related ORF appears to have no particular function in the *Serratia* system.

Table 4-3: Approximation of function of ORFs in the *sea40-sea41* region based on BlastX hits.

Original annotation	New annotation	Query Cover	E value	Per. Ident	Accession
sea40	mRNA interferase HigB [*Serratia symbiotica*]	99%	8.00E-65	95.05%	CDS57329.1
sea41	type II toxin-antitoxin system HigA family antitoxin [*Salmonella enterica*]	99%	4.00E-62	63.27%	EAB5051308.1
-	hypothetical protein [*Salmonella enterica*]	27%	0.000001	93%	EBR7620815.1
-	group II intron reverse transcriptase/maturase [*Salmonella enterica* subsp. Diarizonae]	99%	0	84%	ECG8656393.1
-	hypothetical protein [*Serratia entomophila*]	26%	1E-62	100%	WP_010895838.1
sea42	MULTISPECIES: hypothetical protein [*Serratia*]	99%	6E-94	90%	WP_115185053.1

Table 4-4: Approximation of function of ORFs in the *sea42-sea44* region based on BlastX hits.

Original annotation	New annotation	Query Cover	E value	Identity	Accession
sea42	MULTISPECIES: hypothetical protein [*Serratia*]	99%	6E-94	90%	WP_115185053.1
sea43	hypothetical protein [*Serratia entomophila*]	98%	5.00E-49	100.00%	WP_073999972.1
-	MULTISPECIES: DUF945 domain-containing protein [*Yersinia*]	30%	5.3	48.08%	WP_010895855.1
sea44	DUF945 domain-containing protein [*Serratia entomophila*]	99%	2.00E-139	100.00%	WP_010895855.1
-	transposase [*Pectobacterium polaris*]	98%	3.00E-56	93.48%	KFX15891.1
-	IS3 family transposase [*Erwinia* sp. OLMDLW33]	99%	0	98.59%	WP_099755280.1

Table 4-5: Approximation of function of ORFs in the sea43-sea45 region based on BlastX hits.

Original annotation	New annotation	Query Cover	E value	Identity	Accession
sea43	hypothetical protein [Serratia *entomophila*]	98%	5.00E-49	100.00%	WP_010895854.1
-	MULTISPECIES: DUF945 domain-containing protein [*Yersinia*]	30%	5.3	48.08%	WP_010895855.1
sea44	DUF945 domain-containing protein [*Serratia entomophila*]	99%	2.00E-139	100.00%	WP_010895855.1
-	transposase [*Pectobacterium polaris*]	98%	3.00E-56	93.48%	KFX15891.1
-	IS3 family transposase [*Erwinia* sp. OLMDLW33]	99%	0	98.59%	WP_099755280.1
sea45	autotransporter adhesin Ag43 [*Escherichia coli*]	98%	3.00E-07	27.27%	WP_001825652.1

4.8.2 Insertion between *sea42* and *sea43*

In the isolates 10novel, 25E, E and G the pADAP backbone sequences had a 1946 bp insertion located between the *sea42* and *sea43* ORF, encoding a Domain of Unknown Function (DUF) containing protein encoding gene (Figure 4-10, Table 4-4). These three *sea* ORFs have no known function. With BLAST results returning with hypothetical and transposon hits. The *sea43* and *sea44* ORFs were also anotated as one single ORF in other plasmids analyzed in this study, that were annotated using more current annotation pipelines, and are therefore is most likely one single ORF.

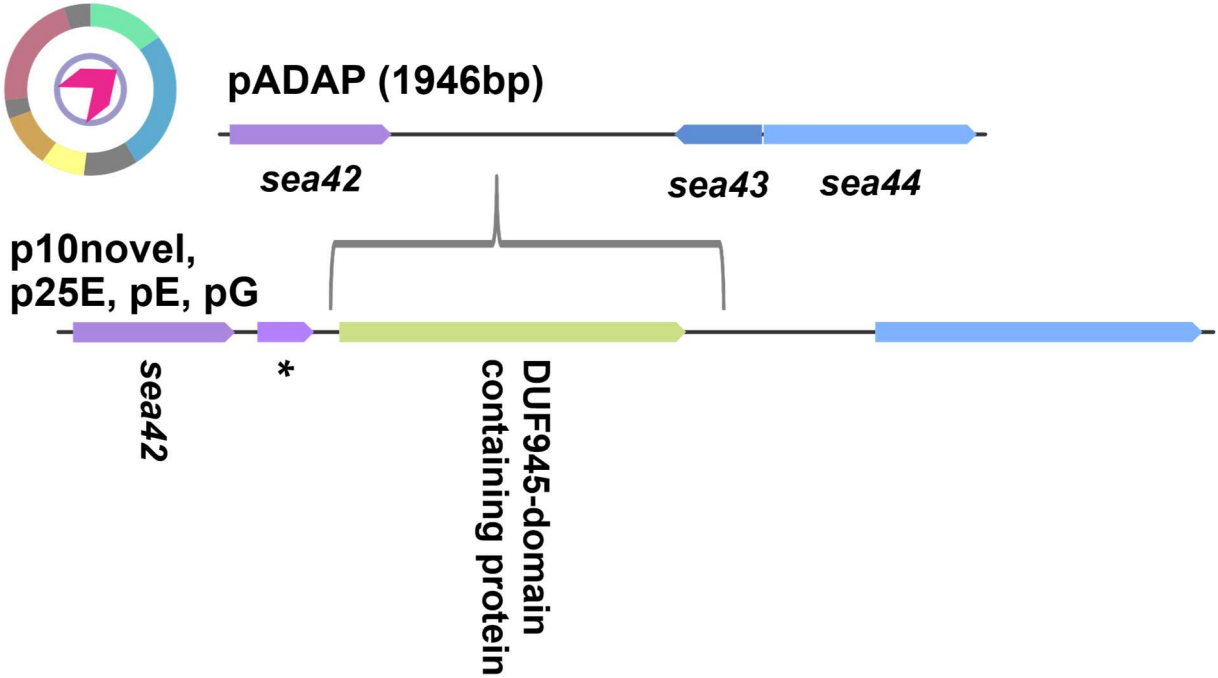


Figure 4-10: The *sea42-44* region in p10novel, p25E, pE and pG. The pADAP backbone sequences derived from the isolates 10novel, 25E, E and G, contain an inserted ORF with a domain of unknown function, a hypothetical ORF (denoted with a *) as well as *sea43* and *sea44* being annotated as a single ORF using the new annotation pipeline.

4.8.3 Insertion between *sea44* and *sea45*

The *sea44–45* region on the RM5 isolate's pADAP plasmid contains a 1663 bp insertion. The inserted region in the pRM5 encodes two putative functional transposable genes (Figure 4-11, Table 4-5). The region itself has no apparent effect on virulence. The transposons are seemingly intact, and therefore most likely active.

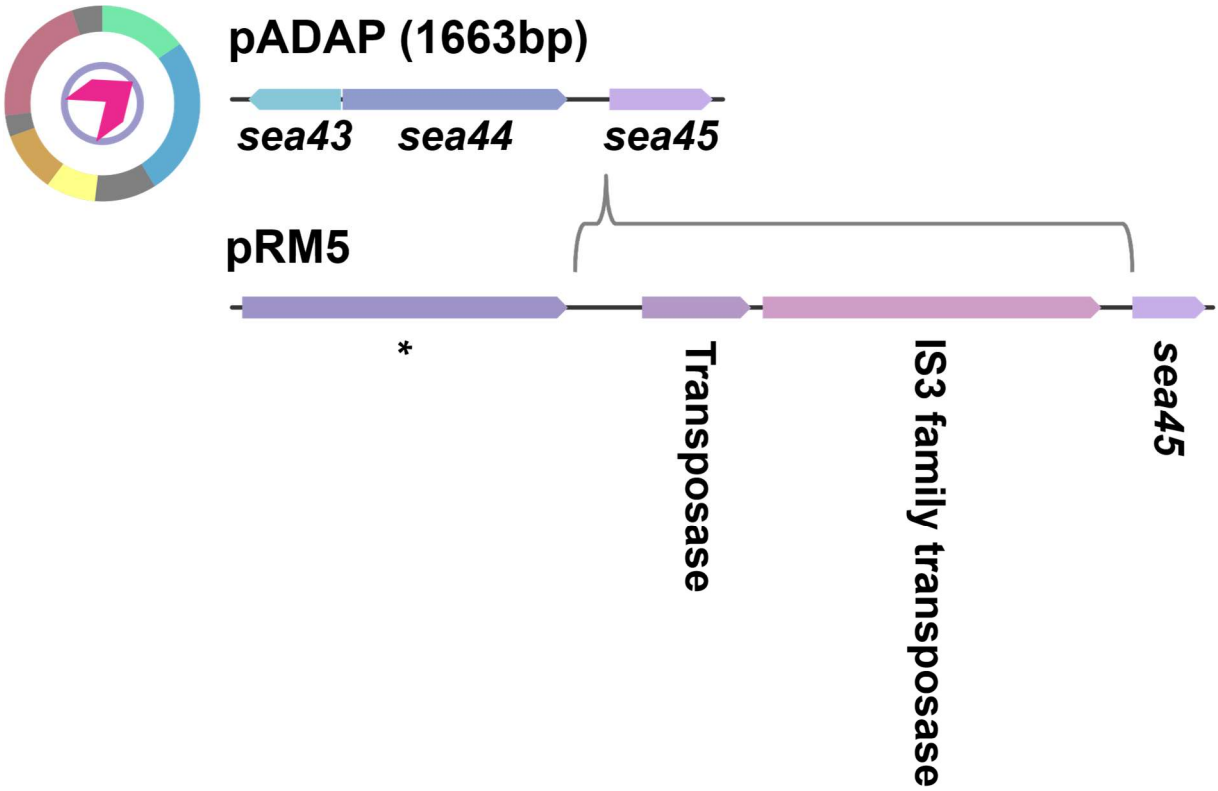


Figure 4-11: The *sea43-45* region in pRM5. The backbone sequence from the RM5 pADAP plasmid contains a DNA insertion of two transposases. These are generic mobile genetic elements and other than transposing themselves in genetic regions, they have no known function.

4.9 Backbone insertion separating *S. entomophila* pADAP from other pADAP

A key finding was the identification of a ~3.9 Kb insertion between the *traG* and *trbC* gene in most plasmids derived from *S. proteamaculans* but not in any *S. entomophila* hosted pADAP plasmids as well as in two *S. proteamaculans* isolates that harbour pADAP orthologs and the US derived WVU-005-1 plasmid (Figure 4-12, Table 4-6). The *traG*, *trbC* and *trbB* genes are part of the Tra/Trb region. Tra is short for transfer, the translated product of *tra* genes are important for relaxing and linearizing the plasmid, priming it for its transfer through the conjugative pore (169, 170). The ~3.9 Kb insertion contains one ORF of interest, encoding for a serine recombinase (Table 4-7). Serine recombinases help with resolving plasmid dimers during replication (306) but also facilitate recombination between two plasmids containing a similar recombination sites (307-309). For cells containing only a pADAP-type plasmid, it is unlikely these recombinases serve any function other than resolution of plasmid dimers, as the pADAP is either single copy or handcuffed prior to cell-division (Section 3.5), thus negating the likelihood of recombination with variants. Enhanced recombination might be possible in the presence of additional non-pADAP plasmids sharing similar recombination recognition sites. Of note, the region contains a ~1.2 Kb intergenic region, which does not encode for any ORFs, however might bind non-coding RNA (ncRNA), although no apparent known binding sites or repeat motifs were identified in this region to confirm this.

This ~5.8 Kb acquired region was absent from all of the assessed *S. entomophila* pADAP plasmids. This region was also found to be absent in the pADAP plasmids obtained from *S. proteamaculans* isolates 142 and 145. This indicates that p142 and p145 plasmids are pADAP ortholog plasmids encoding Sep region (Chapter 5) and Afp orthologs (Chapter 7). The WVU-005-1 plasmid shares the type (A) in respect to Figure 4-12. As this WVU-005-1 plasmid is obtained from another continent, it supports the theory that the region was not lost from the pADAP backbone, but instead was gained in the backbone of pADAP-type variants found in *S. proteamaculans* and *S. liquefaciens* in a more recent insertion event, and could serve as a time stamp, allowing further refinement of the origin of pADAP.

Table 4-6: TraG, TrbC, TrbB insertion type of per isolate.

Type[1]	Species		Isolate
A	*Serratia*	*entomophila*	158, 176, 210, 345, 398, 626, 1100, Diarr, Moraki_2, pADAP
		proteamaculans	142, 145, 1137, 10novel, 25E, E, G, Puna18
		marcescens	WVU-005-1
B		*proteamaculans*	4, 143, 149, 163, 336, 465, 1048, 1071, 1129, 1457, 1769, 1770, 1772, 12a, 12d, 12newD, 1A, 20093, 28F, CfB, D, SpF, K, LC, M, AGR96X, R10, Sprot5
		liquefaciens	376, ,377
C		*proteamaculans*	1, RM5

[1] Type shown in Figure 4-12

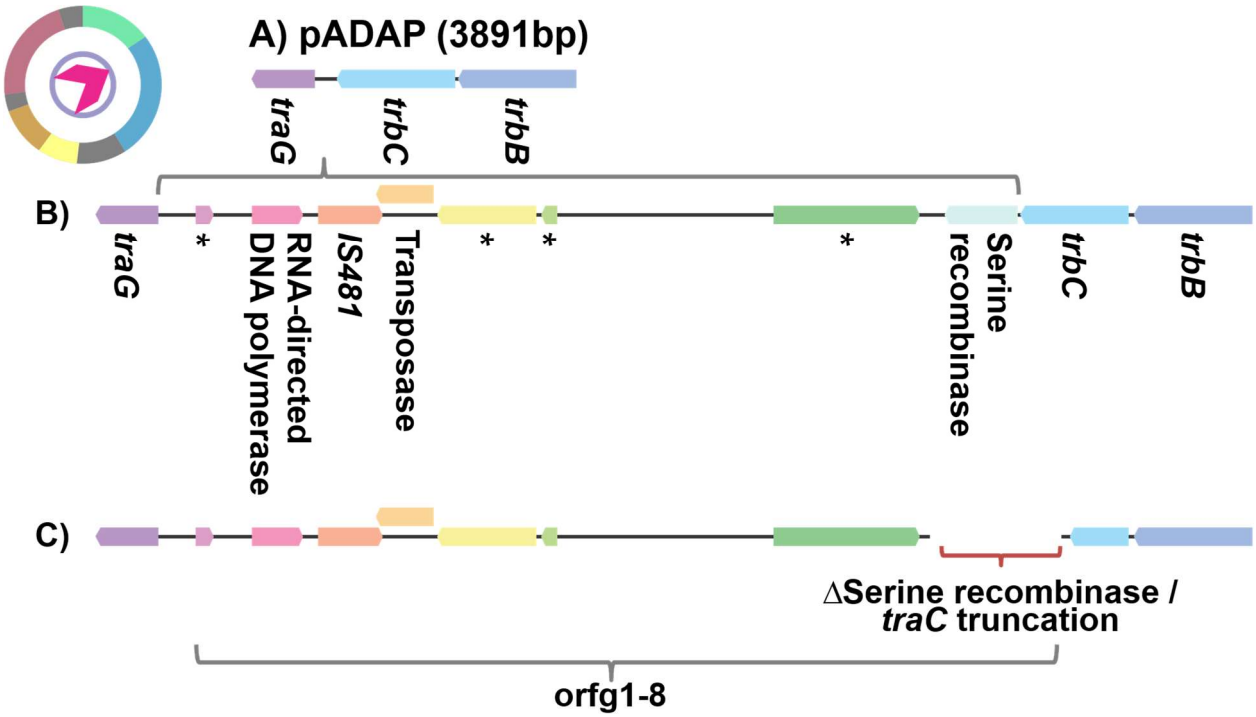


Figure 4-12: The *traG*, *trbC* and *trbB* region that splits the *S. entomophila* pADAP plasmids from the *S. proteamaculans* pADAP-type plasmid variants. All chronic disease inducing S. entomophila pADAP sequences, as well as the *S. proteamaculans* p142 and p145 sequences, share the same three genes (Type A). But most other pADAP plasmids have an ~3.9 Kb inserted region containing two MGEs a polymerase, several hypothetical genes and a Serine recombinase. * denotes hypotheticals (Types B, C).

These regions of insertion are great examples of evolutionary divergence between all plasmids. There are some clear instances where individual isolates have acquired novel elements in backbone regions of the pADAP such as the two heavily divergent regions that are seen in the *sea36-areA* region (Section 4.7) and the *traG*, *trbC* and *trbB* region (Section 4.9), that clearly divide the dataset in smaller subsets. Apart from the shorter regions of divergence described in the previous sections, there are also two examples of large portions of DNA sequence dissimilarity that will be discussed in the following section.

Table 4-7: Approximation of function of ORFs in the *TraG, TrbC, TrbB* region based on BlastX hits.

Gene	Hit							Accession	
traG	MULTISPECIES: phospholipase D family protein [*Serratia*]							WP_115185042.1	
	Length:		178	Range:	1 to 178	Eval:	5E-121		
	Ident:	168/178(94%)		Pos:	173/178(97%)	Gap:	0/178(0%)	Coverage:	0.9438
orfg1	No hits								
orfg2	RNA-directed DNA polymerase [*Vibrionales* bacterium SWAT-3]							EDK27182.1	
	Length:		172	Range:	29 to 103	Eval:	2E-13		
	Ident:	38/77(49%)		Pos:	51/77(66%)	Gap:	2/77(2%)	Coverage:	0.4935
orfg3	IS481 family transposase [*Erwinia tracheiphila*]							WP_046371899.1	
	Length:		346	Range:	1 to 145	Eval:	4E-64		
	Ident:	109/152(72%)		Pos:	117/152(76%)	Gap:	14/152(9%)	Coverage:	0.7171
orfg4	transposase (fragment) [*Xenorhabdus bovienii* str. kraussei Becker Underwood]							CDH24895.1	
	Length:		211	Range:	137 to 201	Eval:	2E-27		
	Ident:	53/66(80%)		Pos:	56/66(84%)	Gap:	1/66(1%)	Coverage:	0.803
orfg5	hypothetical protein [*Serratia marcescens*]							WP_139159366.1	
	Length:		533	Range:	46 to 248	Eval:	3E-109		
	Ident:	168/203(83%)		Pos:	176/203(86%)	Gap:	5/203(2%)	Coverage:	0.8276
orfg6	hypothetical protein [*Serratia marcescens*]							WP_139159366.1	
	Length:		533	Range:	1 to 33	Eval:	8E-10		
	Ident:	29/33(88%)		Pos:	30/33(90%)	Gap:	0/33(0%)	Coverage:	0.8788
orfg7	MULTISPECIES: hypothetical protein [*Serratia*]							WP_013812557.1	
	Length:		313	Range:	1 to 312	Eval:	5E-156		
	Ident:	227/314(72%)		Pos:	263/314(83%)	Gap:	3/314(0%)	Coverage:	0.7229
orfg8	serine recombinase [*Xenorhabdus bovienii*]							WP_038245689.1	
	Length:		207	Range:	112 to 207	Eval:	5E-40		
	Ident:	73/97(75%)		Pos:	80/97(82%)	Gap:	1/97(1%)	Coverage:	0.7526

trbB	thioredoxin fold domain-containing protein [*Yersinia bercovieri*]						WP_145592828.1	
	Length:	333	Range:	1 to 331	Eval:	2E-104		
	Ident:	169/353(48%)	Pos:	219/353(62%)	Gap:	45/353(12%)	Coverage:	0.4788
trbC	MULTISPECIES: conjugal transfer protein TrbC [*Serratia*]						WP_127147021.1	
	Length:	685	Range:	2 to 685	Eval:	0		
	Ident:	667/684(98%)	Pos:	678/684(99%)	Gap:	0/684(0%)	Coverage:	0.9751

4.10 **Conjugation region variants**

As outlined in the previous sections, there is a lot of variation between pADAP backbones. Most of the variation in DNA sequence between pADAP type backbones can be attributed to the insertions of small regions. There are two pADAP-type plasmids however, that have larger variations, most notably in the region transcribing the conjugative machinery (Figure 4-13).

The *S. entomophila* based pADAP plasmids have the region depicted as “pADAP” in Figure 4-13, where most non chronic disease inducing, non-*S. entomophila* pADAP plasmids have the insertion mentioned in Section 4.9, resulting in a region similar to the one depicted as “pAGR96X” in Figure 4-13. The two isolates with very distinct backbones are 465 and 1769. Isolate 465 has an insertion of a potentially intact bacteriophage, between *pilO* and *pilN*, complete with genes encoding for head, tail, and DNA packaging. Of interest, the same ~37 Kb bacteriophage encoding region is also found on a non-pADAP plasmid present in isolate 591, sharing ~88% DNA sequence similarity. This bacteriophage will be further discussed in Section 8.4.

Isolate 1769 is missing a ~23 Kb region, comprising everything from *traT* until the end of the Pil cluster of genes. The last conserved ORF in the Tra cluster is the *traU* ORF, after which the region goes directly into the demarcation region discussed in the Section 4.12. A full manually curated list of deviating ORFs of all pADAP-type plasmid backbones is provided in Supplementary Materials S.40.

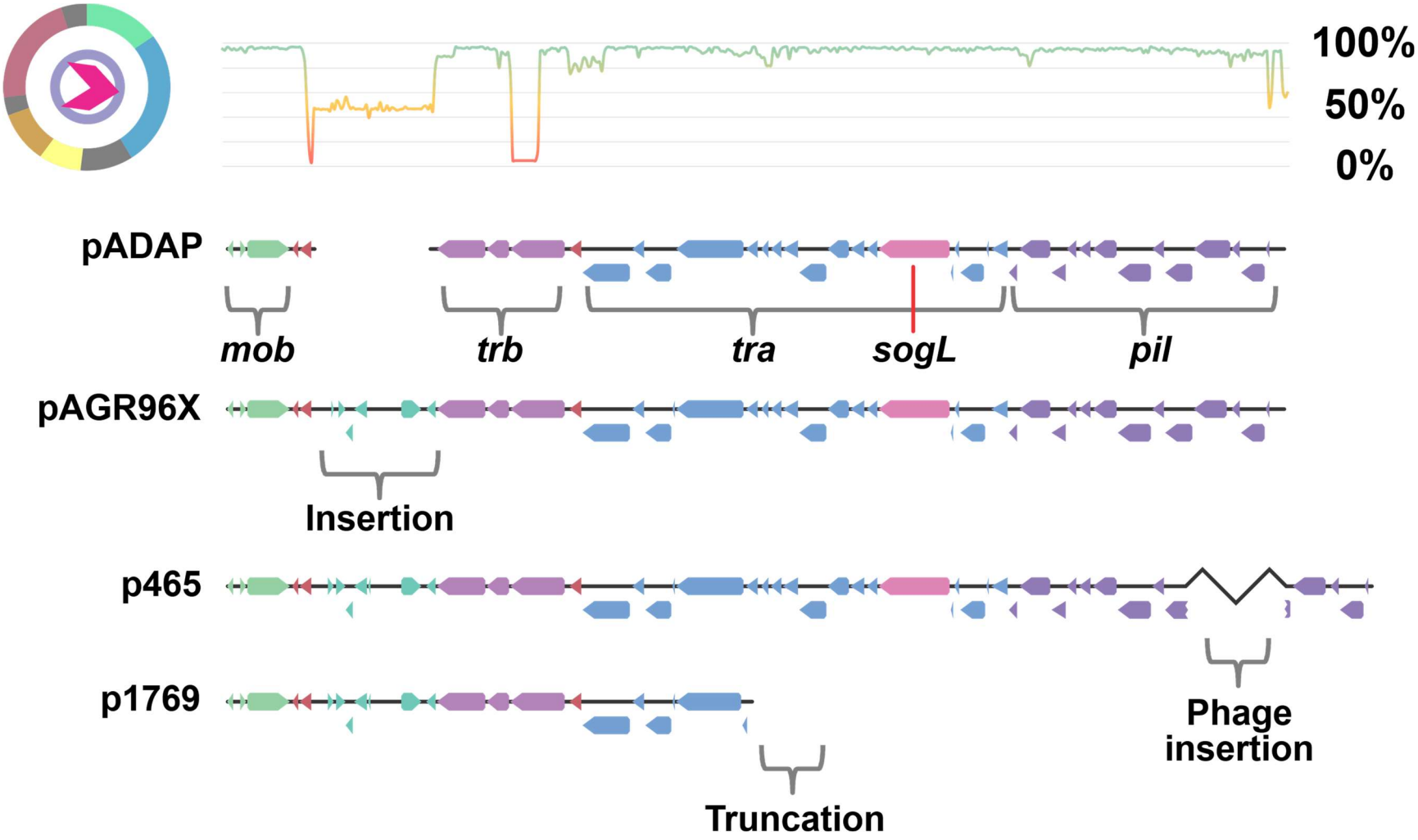


Figure 4-13: An overview of the conjugation region nucleotide sequence variants. Most *S. entomophila* pADAP plasmids, as well the pADAP plasmids found in isolate 142 and 145 and the WVU-005-1 *S. marcescens* plasmid, lack the indicated ~3.9 Kb insertion, previously described in Section 4.9. The plasmid in Isolate 465 also has acquired a bacteriophage encoding region. Isolate 1769 has a ~20 Kb truncation of most of the *tra* and *pil* ORFs in the plasmid. Top part of the figure represents a homology graph of aligned sequences sharing of all pADAP variants sharing this region.

4.11 **Non-New Zealand isolated pADAP-like backbone**

At the latter stages of this study, the *S. marcescens* WVU-005 dataset, containing two plasmids, one of which was designated WVU-005-1, was released. *In silico* analysis revealed that the WVU-005-1 plasmid contained most of the ORFs present in the pADAP backbone. The backbone of WVU-005-1, shares ~80% nucleotide identity to that of the *S. entomophila* pADAP plasmid (Figure 4-7). However, this is an average over the entire backbone. When looking at the actual alignment with the reference pADAP, there are some distinct areas of substantial DNA sequence dissimilarity (Figure 4-14). The most notable region of dissimilarity is from the point of delimitation to the second *par* gene (third on WVU-005-1). The WVU-005-1 plasmid does not encode for a proper ortholog for int2 and repA, nor does it encode the point of delimitation previously described in Section 4.4. The replication genes of WVU-005-1 are located downstream of the backbone, in a region deemed the demarcation region, as this is where most plasmids start deviating substantially. The demarcation region of WVU-005-1 is marked as (I) in Figure 4-15. The WVU-005-1 plasmid does not encode for Afp or Sep, instead, the plasmid is significantly smaller than pADAP, only 91,252 bp in size, and encodes some transposons and secondary metabolites as later outlined in Section 8.6. The second plasmid carried by WVU-005 is WVU-005-2. The WVU-005-2 plasmid is 63,265 bp in size and encodes several Tra and Trb family genes, that have no significant AA sequence similarity with any found on WVU-005-1 or other pADAP-type plasmids.

The *S. marcescens* isolate, containing WVU-005-1, was sourced from a clinical sample from West Virginia, USA. This finding reveals that the backbone of pADAP-like plasmids is more widespread than just New Zealand. Based on the conserved nature of the backbone, it is plausible that this backbone is a generic capturing device that co-evolved with the *Serratia* genus. The differences between the WVU-005-1 plasmid and pADAP is that the pADAP-type plasmids have apparently acquired a novel replication region that is self-incompatible. This is a great example of co-evolution with the host, where a generalist plasmid lost the ability to easily recombine with other pADAP-type plasmids resulting in speciation in the direction of specialization through natural selection.

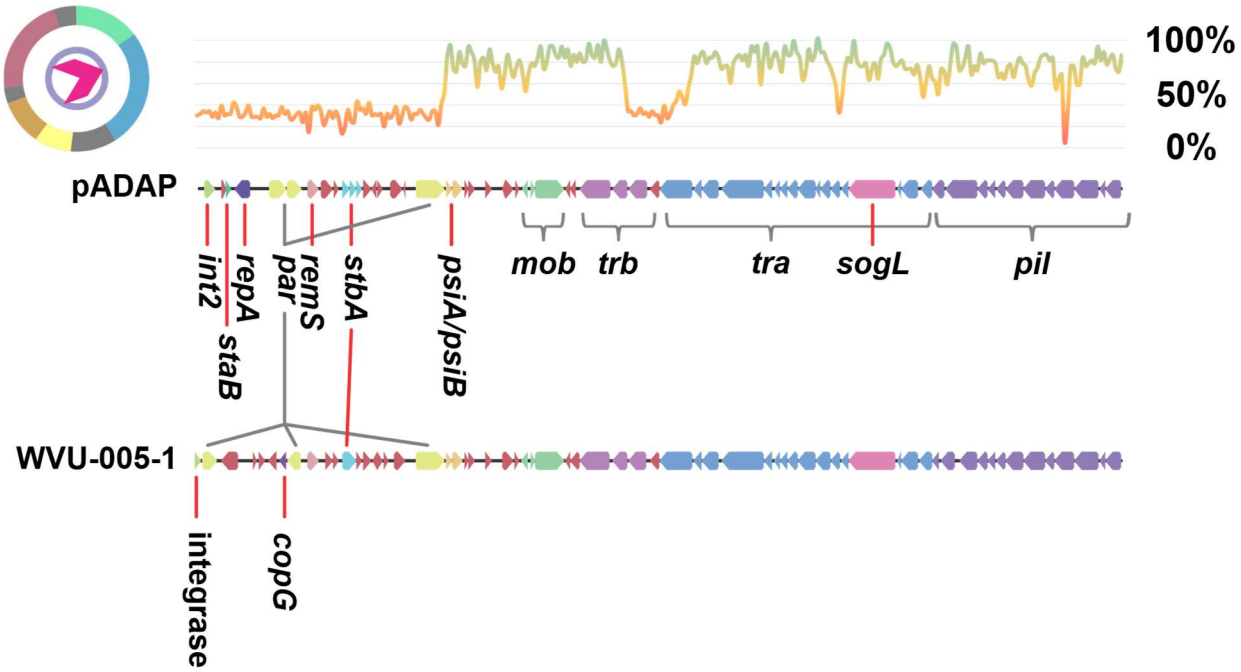


Figure 4-14: Sequence comparison of pADAP [NC002523] and WVU-005-1 [CP041127].

4.12 **Region of low selective pressure**

Historically it was thought that insecticidal pADAP plasmids only contained a Sep and Afp island (190), and there was no reason to believe any other type of island to be present in other isolates in the AgResearch strain library. Most analysed plasmids had either pathogenic or non-pathogenic phenotype, were shown to be similar sizes to pADAP and molecular screening methods such as PCR often confirmed the presence of either the Sep region (161) or Afp region (161). When the sequence of pADAP was published, it was shown that the pADAP variant pU143 (83) was missing the Sep island.

The characteristics of the AGR96X also revealed significant variance in the islands themselves (69). This project further shows that some plasmid variants are missing the Sep islands, others the Afp islands, some do not have either island but instead have completely different islands encoding a diverse range of genes including TAs, phages, etc as outlined in Chapter 8. At the 3' side of the backbone there is a point that was previously assumed to potentially harbour a capturing mechanism that facilitated acquisition of these novel genetic regions and PAI variants by the backbone. However, the region itself, termed the demarcation region, is highly variable, with several non-conserved elements found downstream of the *sea1-sea3* ORFs (Figure 4-15, Table 4-8).

The demarcation region in pADAP is approximately 17 Kb and starts with a cluster of three *sea* genes. The *sea1* and *sea2* ORF have no known ortholog in the RefSeq database and come back as

hypothetical proteins. The *sea3* ORF is identical to a NusG antiterminator. The NusG antiterminator proteins help with transcription of longer regions by reducing the RNA-polymerase half-life (310), and preventing early transcription termination through backtracking and hairpin-stabilized pausing of RNA Polymerases (RNAP) (311). This demarcation region is flanked downstream by the accessory clusters, whether it be a Sef/Sep or Afp region or a long region of nitrogen related genes as is the case in pPuna18, and therefore the NusG may be required for proper transcription of such large operons. All pADAP plasmids present in *S. entomophila* isolates analysed to date share the (A) type of demarcation area, followed by the Sef fimbria region but in the *S. proteamaculans* there is no clear correlation between the elements present in this small island and the type of island that follows as can be seen in Table 4-9.

The predicted function of the IS elements depicted in Figure 4-15 are described in Chapter 9. Isolate 28F is the only pADAP-type plasmid that has the (D) island that is missing *sea3* and *sea4,* in its place there is a region containing a nuclease, domain of unknown function (DUF) 535 containing ORF and two macrolide ORFs. These macrolide ORFs are putative TA proteins. More on these and several other TAs can be found in Chapter 11. The recently sourced plasmid WVU-005-1 does not have the *sea1* nor *sea2* ORF, instead only has a low homology sea3/*nusG* ortholog, followed by two replication genes and a hypothetical gene. It appears this island does not in fact encode any capturing apparatus, and instead could have been acquired through recombination prior to acquisition of other identified accessory clusters. Insertion of transposons and other MGEs as well as the high sequence and gene divergence indicates that this region is under very low selection pressure.

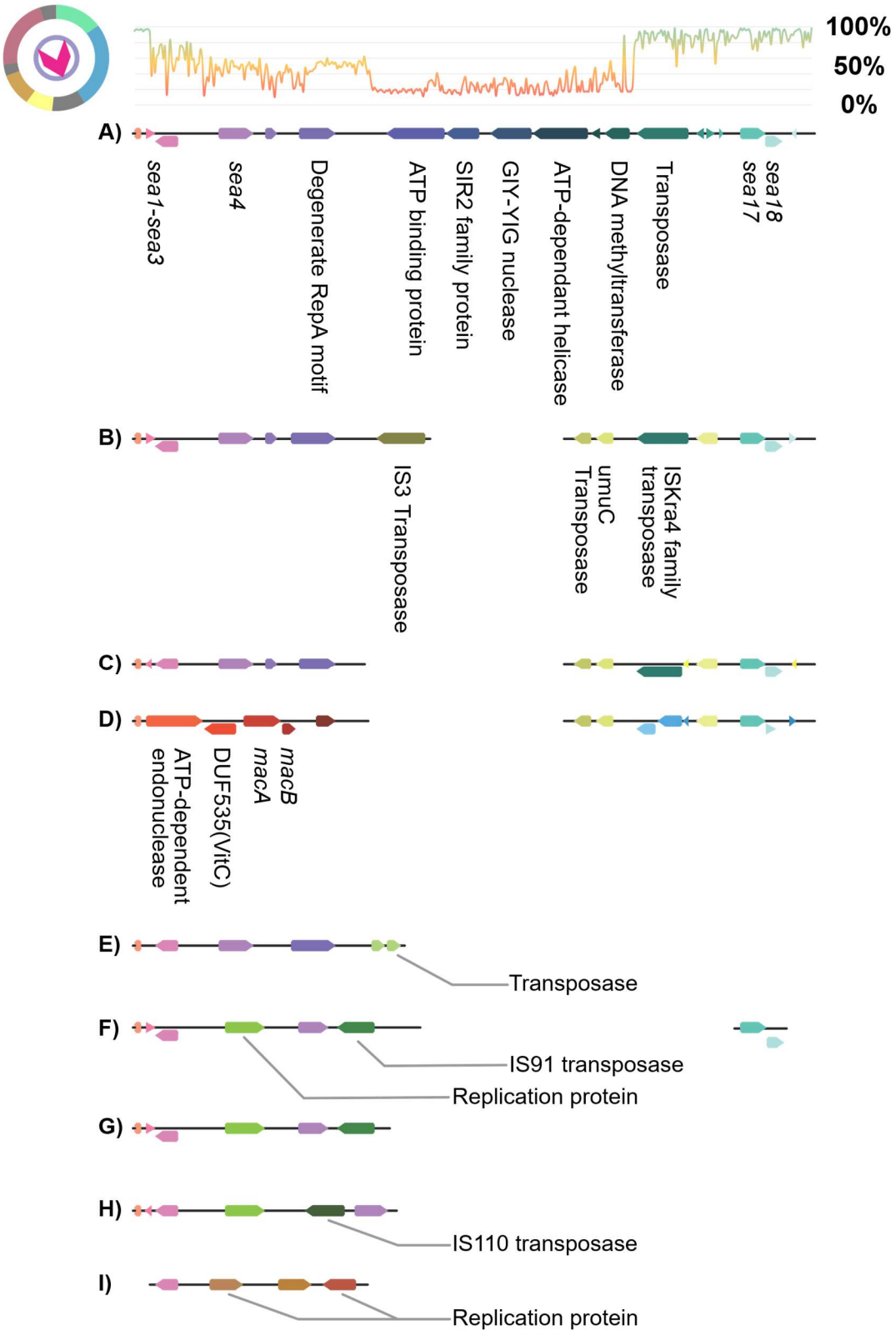


Figure 4-15: Region of low selective pressure, garnering higher rates of mutation. The pADAP type-strain plasmid and all chronic *S. entomophila* pADAP plasmids contains the ~17 Kb region demarcated with A), whereas the other plasmids contain a wide variety of different ORFs. There is no correlation between MGEs and the island downstream of the region. ORFs that are not annotated are hypothetical, unknown genes or genes containing a Domain of Unknown Function (DUF)

Table 4-8: Demarcation variant type per isolate as shown in Figure 4-15.

Type	Species		Isolate
A	*Serratia*	*entomophila*	158,176, 210, 345, 398, 626, 1100, Diarr, Moraki_2, A1MO2
		proteamaculans	142, 145
B		*proteamaculans*	143, 1071, 10novel, 25E, CfB, D, E, G
C		*proteamaculans*	149, 299, 336, 465, 1048, 1457, 1769, 1770, 1772, 12a, 12d, 12newD, K, M
D		*proteamaculans*	28F
E		*proteamaculans*	Puna18
F		*proteamaculans*	1, 1129, R10, RM5
		liquefaciens	376, 377
G		*proteamaculans*	4, 163, 1A, 20093, SpF, LC, AGR96X, Sprot5
H		*proteamaculans*	1137
I		*marcescens*	WCU-005-1

Table 4-9: Demarcation region types and their association with the Sep and Afp PAIs.

Sef/Sep on plasmid	Afp on plasmid	Demarcation region
+	+	A, F
+	-	F, B, C, D
-	+	G, H
-	-	E, I

Chapter 5 *Serratia entomophila* pathogenicity toxin complex

The Sep proteins belong to the ABC toxin complex (Tc) family. As described by Hurst et al. (83), there are both Sep and Spp variants of this toxin complex. Initially Sep variants were found in *Serratia entomophila* and designated Sep for *S. entomophila* pathogenicity complex. In subsequent years, PAIs were found in *Serratia proteamaculans* with low homology scores to Sep, but with the same composition and function, these Tc orthologues were termed *Serratia proteamaculans* pathogenicity (Spp). These ABC toxin complexes comprise of three components that form a syringe like structure capable of transporting protein based toxins from the extracellular matrix (EM) through the outer membrane (OM), into the periplasm (PP) of the target cell (47) (Figure 1-8).

5.1 Sep/Spp nucleotide similarity

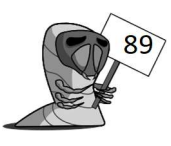
A direct homolog of the Sep operon has previously been identified in the *Yersinia frederiksenii* strain 49 (90), which, like *Serratia*, is also part of the *Yersiniaceae* family (66). This homolog was designated as TcYF. In this study five additional non-pADAP plasmids were identified that have acquired the Spp PAI seemingly independently, which is a good indication of horizontal transfer of the pathogenic element. The fact that the Sep operon is highly similar to the toxin complexes found in *Photorhabdus luminescens* (190, 191, 312), reinforces this suggested shared ancestory through HGT (192). The Sep/Spp islands on pADAP-type plasmids are flanked upstream by the Sef region, that will be described further in Chapter 6, and downstream mostly by either the Afp region that will be discussed in Chapter 7, or a conserved genomic island (GI) (313), designated as Region of Unknown Function (RUF), discussed later in Section 8.1. There are also several Spp on plasmids that have no additional ORFs between the PAI and the backbone.

The non-pADAP encoding Spp PAIs, encoded by isolates S-prot-1, Man4, 591, 49, Man3, D1 and Sm1a, are almost exclusively flanked by a Sef fimbria cluster encoding region upstream. The non-pADAP carrying TcYF PAI found on *Y. frederiksenii* based p49 plasmid however, is the only Sep PAI homolog that has no associated Sef region. All isolates with non-pADAP plasmids encoding for a Spp PAI orthologs, were found to be avirulent through bioassays (Table 5-1). S-prot-1 was not bioassayed but is assumed to have similar bioactivity. This suggests that in those instances, the Spp PAI is not the only element required for bioactivity, and that there might be a regulatory difference between these plasmids.

Table 5-1: Bioassay results of non-pADAP Spp encoding isolates compared to A1MO2.

Strain	Species		Disease%	Mortality%	Affected%
Grass Grub					
Blank			0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
A1M02	*Serratia*	*entomophila*	83.3 ± 4.1	3.6 ± 2.0	86.9 ± 3.7
591		*proteamaculans*	0.0 ± 0.0	8.3 ± 8.3	8.3 ± 8.3
49	*Yersinia*	*frederiksenii*	8.3 ± 5.8	0.0 ± 0.0	8.3 ± 5.8
Man4	*Serratia*	*proteamaculans*	8.3 ± 5.8	12.5 ± 6.9	20.8 ± 8.5
Sm1a		*proteamaculans*	5.0 ± 8.3	0.0 ± 0	5.0 ± 8.3
D1		*entomophila*	12.5 ± 12.5	87.5 ± 12.5	100 ± 0.0

No MGE, such as IS and transposons, nor nucleotide “scars” that can indicate an active site of trans positioning/mobilization of the PAI, were found to be shared among all Sep/Spp operons, however elements were found to be shared within subsets of these Sep/Spp PAIs. MGEs will be discussed later in Chapter 9. The sequence nucleotide similarity matrix of the Sep/Spp region shows that apart from the *S. entomophila* pADAP based *sep* operon, there is very poor sequence conservation over all operons, which correlates with the lack of shared MGE elements or “scars”, across these different acquired Tc encoding islands, revealing a large evolutionary distance between the PAIs (Figure 5-1) and divergence through means other than HGT.



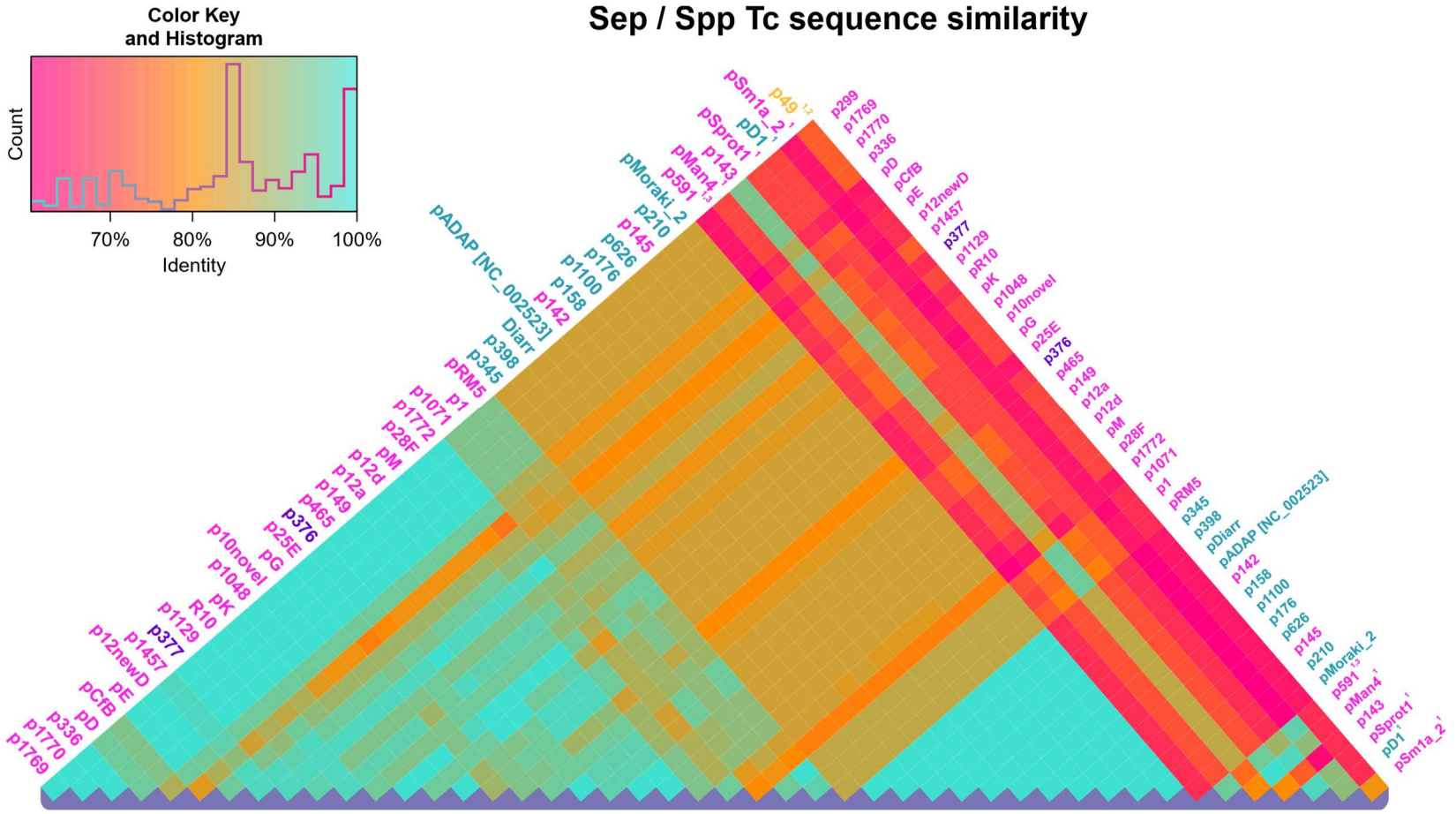


Figure 5-1: A DNA similarity matrix of the Sep/Spp regions obtained from 46 different plasmids, six of which are non-pADAP-type plasmids. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens*, orange *Y. frederiksenii* and pink *S. proteamaculans*.

[1] Non-pADAP-type plasmids.

[2] p49 does not encode Sef region upstream, is avirulent and is the only Yersinia plasmid found with the Sep ortholog TcYF

[3] The non-pADAP-type plasmid that shares the bacteriophage region with the pADAP-type plasmid p465.

5.2 **Sep/Spp alignment features**

The DNA regions in the Sep/Spp PAI that are causal for the low nucleotide sequence similarity between PAIs (Figure 5-1) can be seen more clearly when looking at the sequence alignments and annotations themselves (Figure 5-2, Table 5-2). Full DNA alignment of all the Sep/Spp encoding region can be found in Supplementary Materials S.22. All Sep variants only contain one *sepC* ORF, although annotation variations exist for the *sepA* ORFs. The Spp comprise of several variants with different numbers of *sppC* ORFs as well as annotation variations, truncations, deletions and some cases of absence of a *sppC* ortholog. Apart from variations in the *sppC* there are also annotation variations in *sppA* and *sppB*. The DNA homology graph clearly shows that most regions of low sequence conservation are in the A and C component of these toxin complexes, with the B component being reasonably conserved across variants (>99%). Based on the resolved structure of YenTc (197, 314, 315) and TcdTc (48), this conservation is in line with the structural formation of the Tc, wherein the B component encapsulates the C and connects the B-C to the A as has been shown at the structural level in the YenTc complex (197), which is a Sep ortholog (316). The B component is also implicated in transportation of the C toxin through the SepA induced pore, as demonstrated in the TcdTc ortholog found in *Photorhabdus luminescens* (48) (Figure 1-8).

Several degenerate and orphan components of the Spp complex were found in both pADAP-type plasmids and non-pADAP-type plasmids, as depicted in Figure 5-2. There are two *sepA*/*B* orphans that are not shown in Figure 5-2, as they are part of a highly conserved RUF GI found in several plasmids. This RUF region is further discussed in Section 8.1.

Full DNA alignment of the Sep/Spp regions can be found in Supplementary Materials S.22.

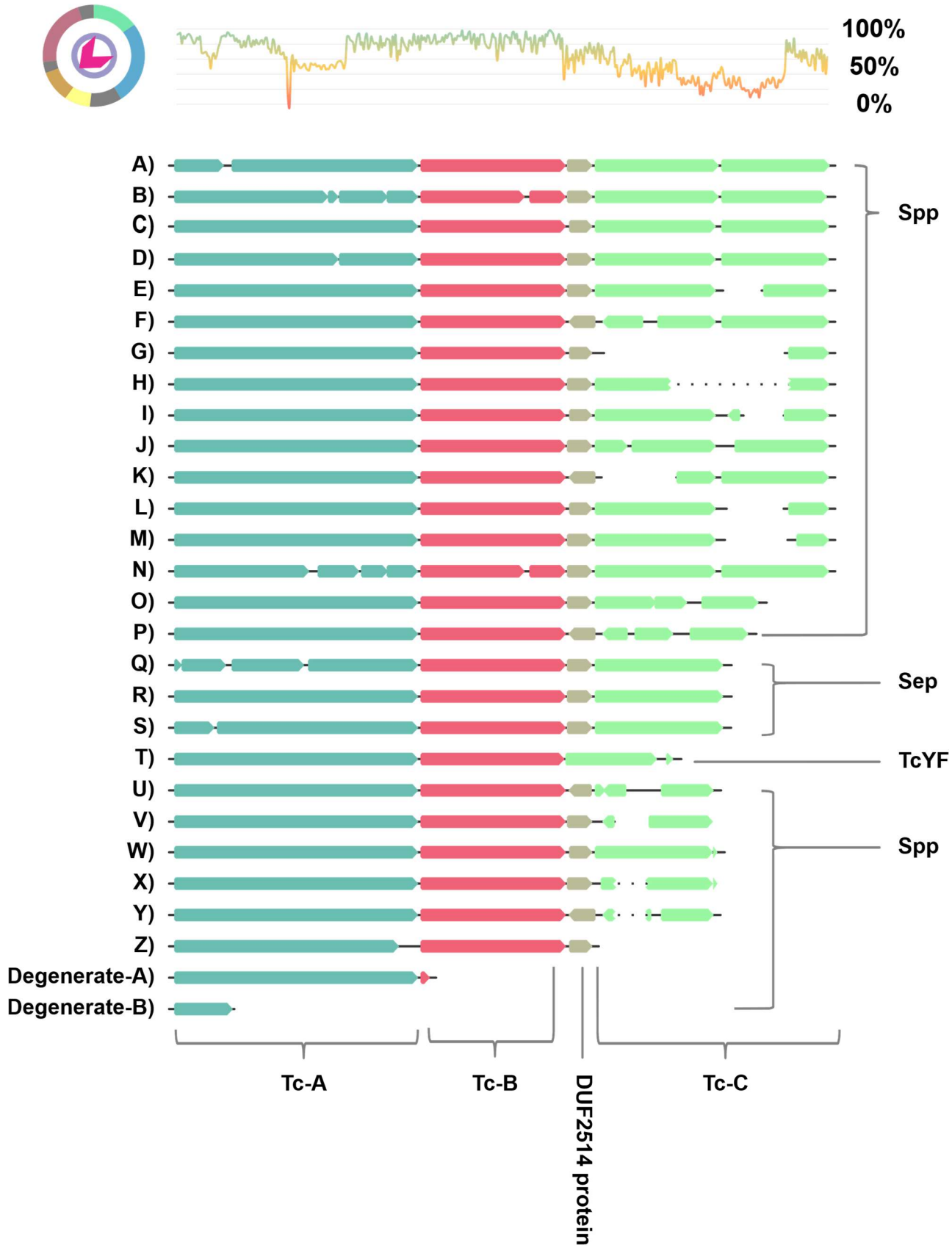
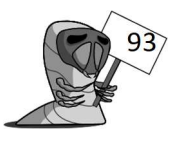


Figure 5-2: DNA alignment of all Sep and Spp orthologous regions.Refer to Table 5-2 for plasmid types A-Z.

Table 5-2: Sep / Spp variant type per isolate. pADAP-type plasmid variants are on the top and non-pADAP-type plasmids at the bottom.

Variant	Species		Type	Isolate
pADAP-type				
C	*Serratia*	*proteamaculans*	Spp	149, 465
D			Spp	RM5
E			Spp	1
F			Spp	1071
G			Spp	1772
H			Spp	28F
I			Spp	12a, 12d, M
J		*liquefaciens*	Spp	376
K		*proteamaculans*	Spp	25E
L			Spp	G
M			Spp	10novel
N			Spp	1048
O			Spp	E
P			Spp	CfB
Q			Sep	142
R		*entomophila*	Sep	145, 158, 176, 210, 398, 626, 1100, Diarr, K, Moraki_2, pADAP
S			Sep	345
U		*proteamaculans*	Spp	143
V			Spp	R10
W			Spp	1129
X		*liquefaciens*	Spp	377
Y		*proteamaculans*	Spp	12newD
Z			Spp	1457
Degenerate A			Spp	299, 336, 1769, 1770
Non-pADAP				
A	*Serratia*	*proteamaculans*	Spp	S-prot-1
B		*proteamaculans*	Spp	Man4
C		*proteamaculans*	Spp	591
T	*Yersinia*	*frederiksenii*	TcYF	49
Degenerate A	*Serratia*	*entomophila*	Spp	Man3, D1
Degenerate B		*proteamaculans*	Spp	Sm1a



The low homology of the Tc encoding islands (Figure 5-1, Figure 5-2) shows that the islands have undergone substantial DNA mutations, suggesting they are most likely one of the first elements acquired by pADAP, as other acquired islands such as the RUF GI (Section 8.1) and Afp PAI (Chapter 7) are conserved to a much higher degree at the nucleotide level. As shown in Figure 5-3, the GC content of the Sep encoding region in pADAP also confirms that the acquisition was not recent, as the GC% is similar to that of the replication region, and slightly higher than that of the rest of the backbone (not shown), indicating amelioration. This is shown for all for all Sep and Spp encoding variants in Figure 12-9. The GC content of PAIs and GIs will be further discussed in Chapter 12 where the genetic diversity of the pADAP family of plasmids as a whole, is further analysed.

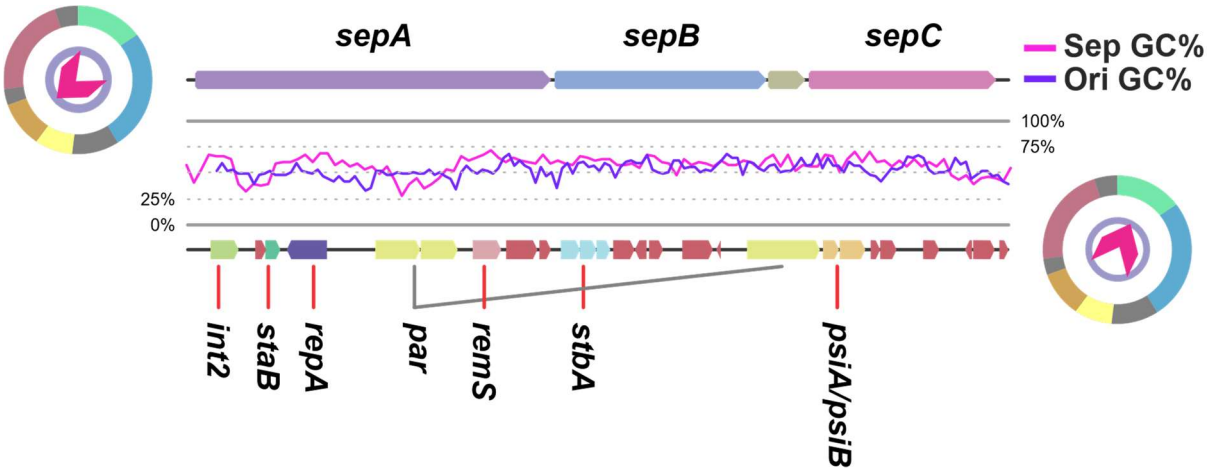


Figure 5-3: Comparison of the GC% of the Sep operon, compared to the replication region of the pADAP plasmid. The Sep operon has a relatively similar, and overall higher GC% than the entirety of the replication region has, indicative of a stable element in the plasmid. The Sep is represented as the top feature map, and its GC% in pink, whereas the replication region is represented as the bottom feature map, with its GC% in purple.

Chapter 6 *Serratia entomophila* fimbriae

Located upstream of Tc cluster (Chapter 5), and downstream of the previously described demarcation region outlined in Section 3.5 (Figure 1-6), the pADAP plasmid encodes for a Type 1 fimbria (83, 183). The region was designated as the Sef operon for *Serratia entomophila* fimbria.

6.1 Sef encoding region

The Sef encoding region encompasses ten ORFs, *sefA*-*sefJ*, as well as the *sea19* ORF (Figure 1-6). For a detailed view of the region as well as functional description of the ORFs and breakdown of types, refer to Figure 6-1, Table 6-1 and Table 6-2 respectively. Across all Sef encoding regions of *Serratia* isolates sequenced through this study, the sef encoding region was found to be highly conserved at the nucleotide level and ORF orientation, as can be seen from the homology graph in Figure 6-1 as well as the from the nucleotide distance matrix (Figure 6-2). Surprisingly the region is also highly conserved in several non-pADAP plasmids that carry the Spp encoding PAI such as pD1, pS-prot-1, p591, pSm1a, pM3 and pMan4, although these do contain the additional genes 3' of the Sef encoding cluster, as depicted in sequence type E, F and G (Figure 6-1). The Sef operon is almost co-located 5' of the Sep or Spp region, but missing from the TcYF cluster, as depicted in Figure 6-3.

Full nucleotide alignment of the Sef regions can be found in Supplementary Materials S.20.

Table 6-1: Sef variant type (see Figure 6-1) per isolate encoded on pADAP-type plasmid and non-pADAP plasmids.

Variant	Species		Isolate
pADAP-type			
A	*Serratia*	*entomophila*	158, 176, 210, 345, 398, 626, 1100, Diarr, Moraki_2, pADAP
		proteamaculans	143, 145, 149, 336, 465, 1071, 1457, 1769, 1772, 12a, 12d, 25E, 28F, CfB, D, G, M, R10
B		*proteamaculans*	1, 299, 1129, 1770, 10novel, 12newD, E, K
		liquefaciens	376, 377
C		*proteamaculans*	142, 1048
D		*proteamaculans*	RM5
Non-pADAP			
E	*Serratia*	*proteamaculans*	D1, S-prot-1
F		*proteamaculans*	591, Sm1a
G		*proteamaculans*	Man4

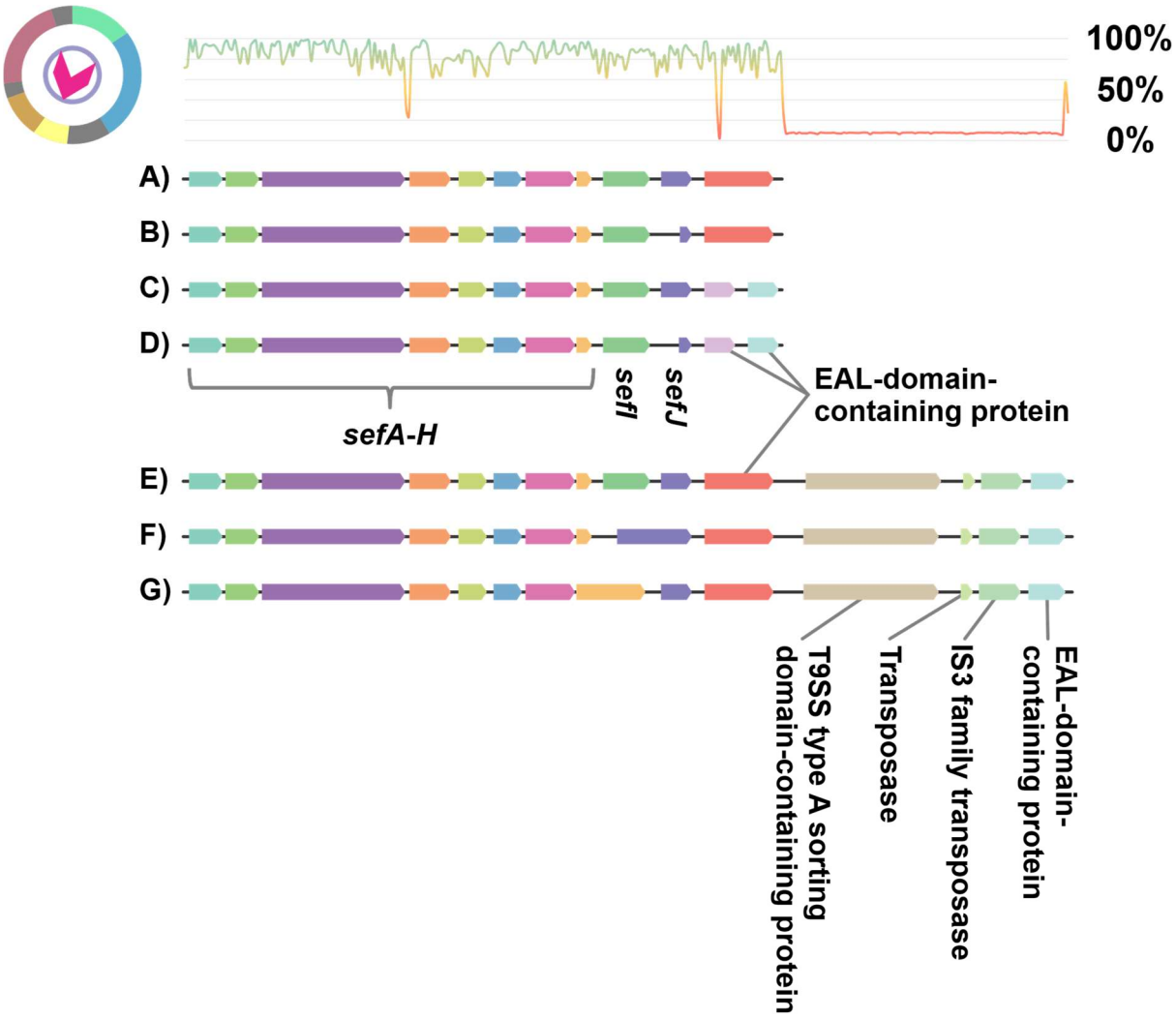


Figure 6-1: Nucleotide alignment of all orthologous Sef regions. This region is co-located 3' to the Spp operon in *Serratia* pADAP-type plasmids and non-pADAP-type plasmids. All non-pADAP-type plasmids were found to encode 4 additional genes, depicted in E, F and G, located between the Sef and Spp operons.

Table 6-2: Approximation of function of ORFs in the Sef region based on BlastX hits.

Type	Name	Size (AA)	Closed functional description	Query Cover	E value	Identity	Accession
A-G	SefA	176	type 1 fimbrial protein [*Aeromonas simiae*]	98%	3.00E-69	64.77%	QFI54445.1
A-G	SefB	176	MULTISPECIES: type 1 fimbrial protein [*Providencia*]	88%	3.00E-41	45.86%	WP_109913232.1
A-G	SefC	844	fimbria/pilus outer membrane usher protein [*Enterobacter ludwigii*]	95%	0	61.14%	WP_163387471.1
A-G	SefD	245	fimbria/pilus periplasmic chaperone [*Erwinia teleogrylli*]	94%	2.00E-112	63.79%	WP_058911439.1
A-G	SefE	208	MULTISPECIES: type 1 fimbrial protein [*Enterobacteriaceae*]	82%	3.00E-54	51.46%	WP_024547671.1
A-G	SefF	174	type 1 fimbrial protein [*Kluyvera intermedia*]	85%	8.00E-44	49.32%	WP_062777744.1
A-G	SefG	263	fimbria/pilus periplasmic chaperone [*Proteus hauseri*]	92%	4.00E-81	46.75%	WP_064719178.1
A-G	SefH	175	MULTISPECIES: fimbrial protein [*Providencia*]	83%	1.00E-44	58.22%	WP_109913237.1
A-E	SefI	271	adhesin [*Proteus columbae*]	41%	8.00E-09	39.32%	WP_100158730.1
A-G	SefJ	198	adhesin [*Proteus columbae*]	61%	6.00E-05	31.45%	WP_100158730.1
A-G	Sea19	409	EAL domain-containing protein [*Hafnia alvei*]	80%	3.00E-115	52.41%	WP_166492956.1
E-G	NA	827	T9SS type A sorting domain-containing protein [*Apibacter adventoris*]	58%	3.00E-32	27.78%	WP_146105740.1
E-G	NA	89	putative tansposase [*Shimwellia blattae* DSM 4481 = NBRC 105725]	98%	5.00E-46	96.59%	AFJ45628.1
E-G	NA	267	IS3 family transposase [*Serratia fonticola*]	99%	0	90.60%	WP_080670645.1
E-G	NA	213	EAL domain-containing protein [*Serratia entomophila*][1]	99%	9.00E-130	96.86%	WP_010895790.1

[1]This gene appears to be homologous to *sea19*

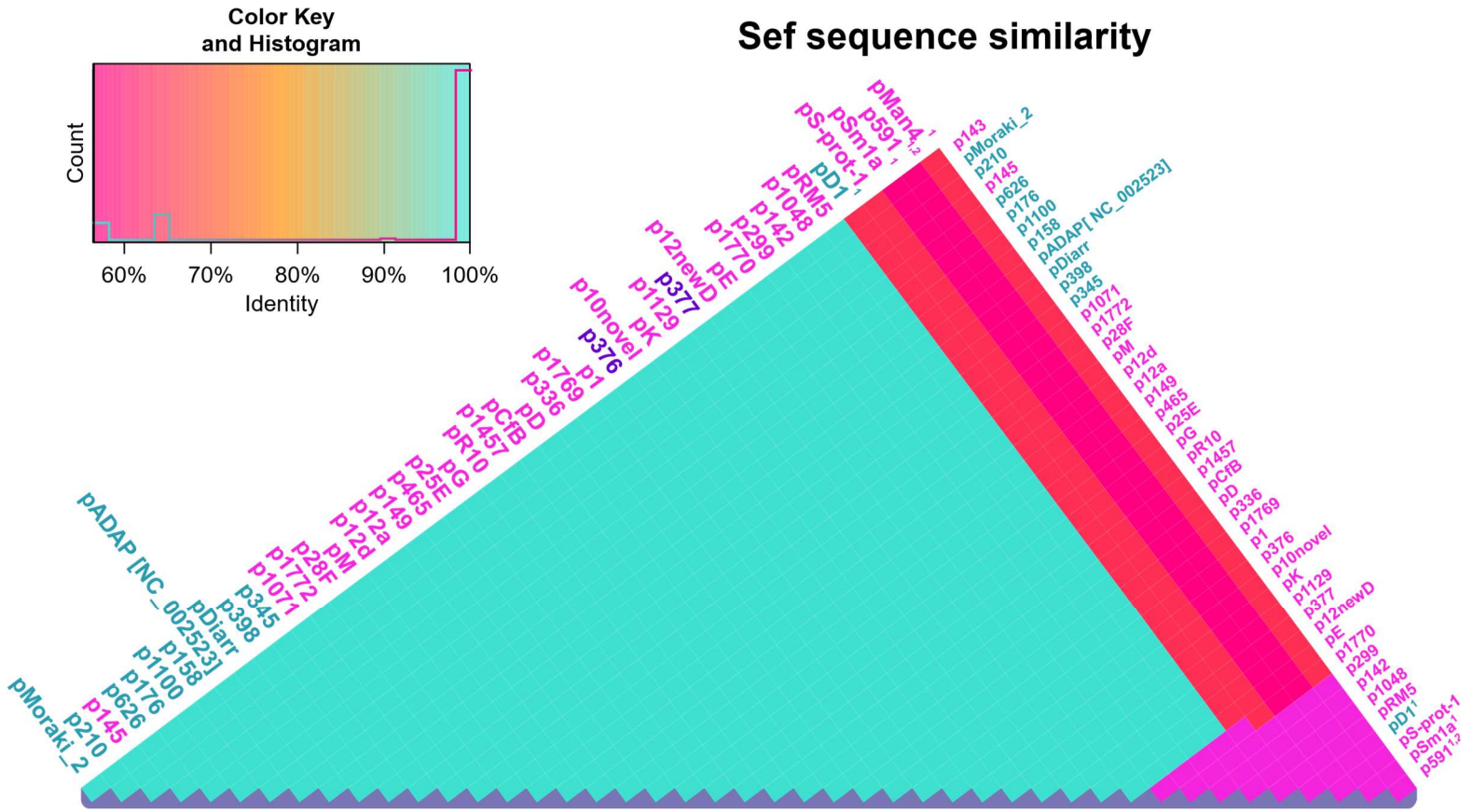


Figure 6-2: A DNA similarity matrix of the Sef encoding regions derived from 45 different plasmids, five of which are non-pADAP-type plasmids. Of note is the *Yersinia frederiksenii* p49 is the only Sep encoding plasmid that does not encode for a Sef region. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens* and pink *S. proteamaculans*.

[1] Non-pADAP-type plasmids.

[2] The non-pADAP-type plasmid that shares the bacteriophage region with the pADAP-type plasmid p465.

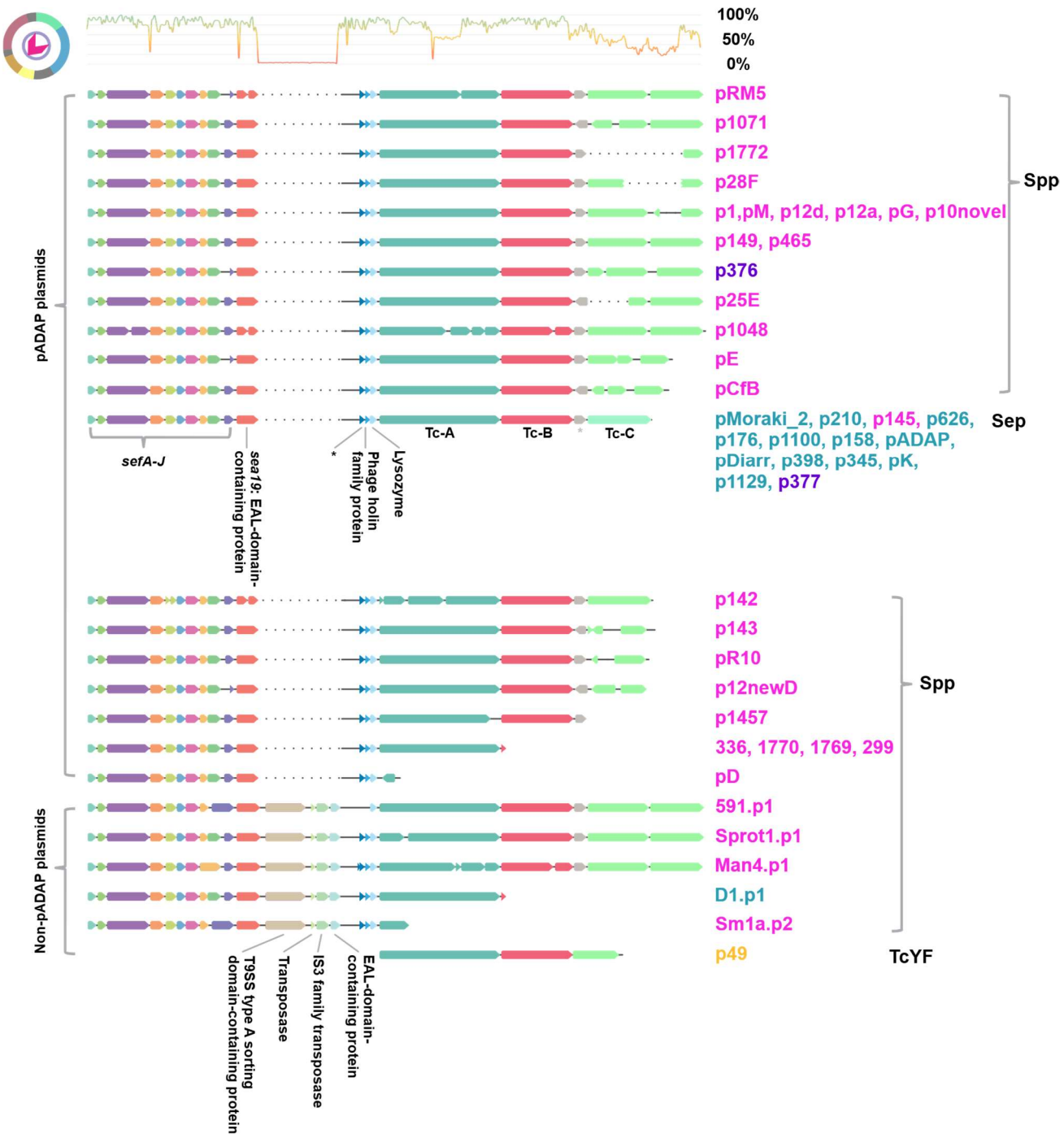


Figure 6-3: Nucleotide alignment of the entire Sep/Sef operon of all plasmids carrying this region. All non-pADAP-type plasmids were found to encode 4 additional genes, depicted in E, F and G, located between the Sef and Spp operons. The *Y. fredrekensii* p49 plasmid, encoding for the TcYF, a Spp ortholog, does not encode Sef. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens*, yellow *Y. frederiksenii*.

6.2 **Characterization of Sef**

Type 1 fimbriae are associated with colonization of numerous types of organisms, from humans (183, 184), mice (185) and swine (186), to invertebrates such as nematodes (188) and insects (188). Most Type 1 fimbriae produced by *E. coli* are encoded for by FimA-H (317) (Table 6-2). A schematic of the Fim complex is depicted in Figure 6-4. In this fimbrial type, FimC is a chaperone that transports all Fim components from the periplasm to the OM and FimC is homologous to SefD. FimD is an usher protein that forms a complex that can transports fimbrial components across OM to the EM. SefC is the Sef equivalent of FimD in function. FimA, a distant ortholog to SefA, is the protein that forms the fimbrial rod and FimF, G and H form the tip fibrillum. FimH is associated with adherence (317). In *E. coli*, the FimH adhesin binds to D-mannose residues found in glycoproteins on various types of cell membranes (318).

Although Sef appears to form a Type 1 fimbria, the nucleotide sequence in its entirety does not have any DNA orthologs in the RefSeq database. When analysing the amino acid sequences, the *sefA-F* ORFs share amino acid similarity to Type 1 fimbriae orthologs, commonly found in *E. coli*. There is no FimH ortholog present in the Sef gene cluster. FimH is the component of the Type-1 fimbria that is needed for adherence to D-mannose residues, found on certain glycoproteins on cell surfaces (319). SefH is orthologous to the fimbrial minor subunit StfF of the Stf fimbriae found in *Salmonella enterica* (320). The Stf fimbria is a poorly described fimbria and not much research has been undertaken on the individual components other than StfA (321). Some *Salmonella*-based fimbriae have been implicated in mammalian gut colonization (322) however, the *Salmonella enterica* subspecies II serovar, from which the StfF ortholog is derived, has only on rare occasions been linked to mammalian infections (323). The Stf fimbria itself has been shown to have a no significant role in intestinal colonization (321). Based on this information, it is not possible to derive functional knowledge about SefH from its closest documented ortholog, StfF. Both SefI and SefJ share low amino acid similarity with adhesins of the *Morganella* and *Proteus* genus (Table 6-3). These two genera, together with the genus *Providencia* that encodes for a fimbrial protein, homologous to SefH, are all gram-negative bacteria, often associated with humans and other mammals and on rare occasions are implicated in mammalian diseases (324). As the role of these orthologous fimbrial proteins and adhesins have not been determined in these bacteria, combined with their limited AA sequence similarity to SefI and SefJ, it can only be assumed that they have a role in adhesion to a yet to be determined substrate

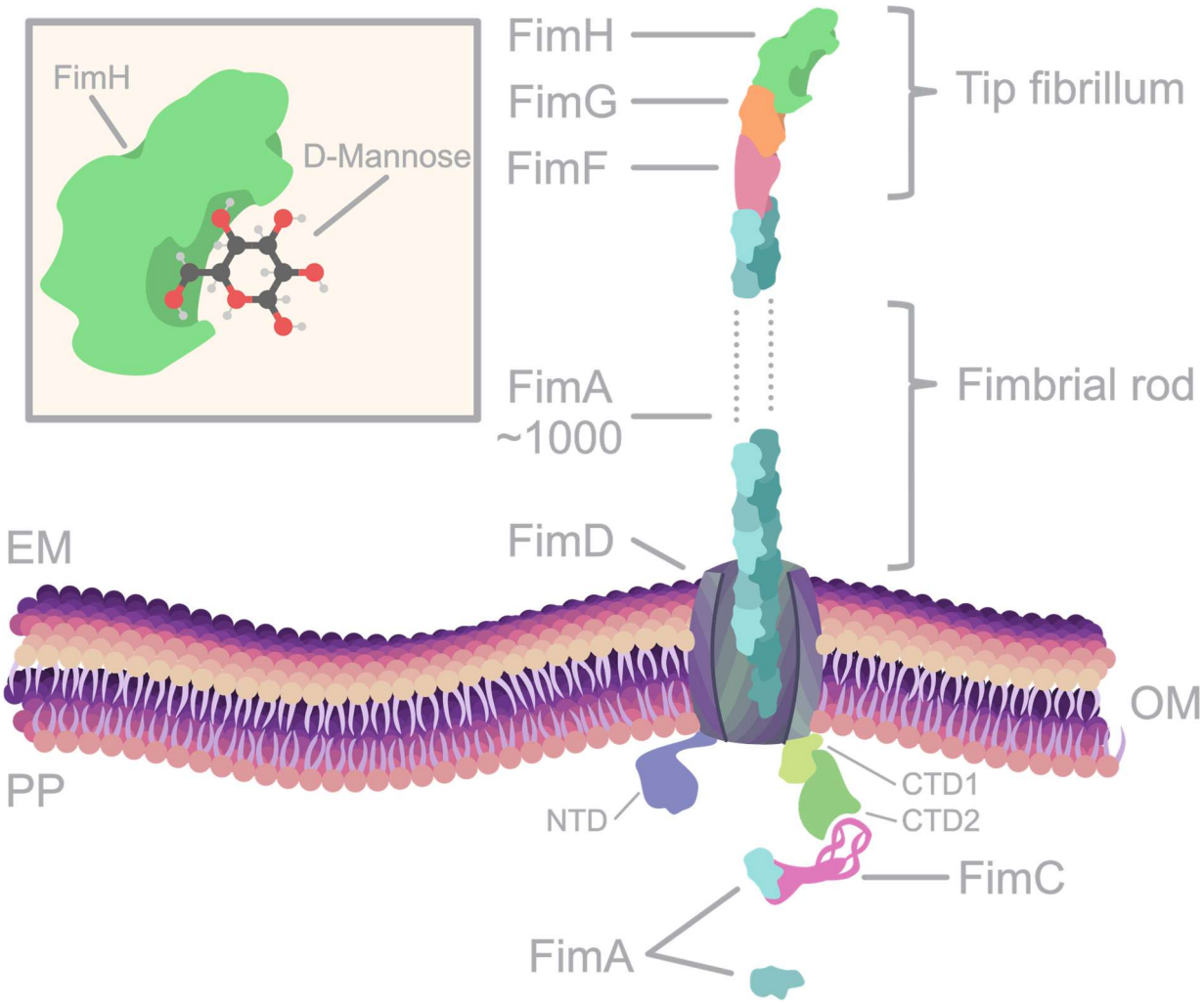
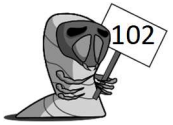


Figure 6-4: A simplified model of the Type 1 fimbriae typically found in *E. coli*. FimC transports the FimH followed by FimG, FimF and FimA proteins from the periplasm to a FimD usher in the OM. The FimD transports the fimbria components through the OM into the EM. The FimH tip has high binding affinity for D-Mannose, a residue found in glycoproteins on OMs. The Sef fimbria consists of mostly Type 1 fimbriae orthologs, although several genes, most notably the *sefH*, *sefII* and *sefJ* appear to be unique. The *sefH-J* ORFs most likely form the tip fibrillum, that adheres to something other than D-Mannose residues. Figure is based on models by Busch et al. (317), Godaly et al. (318), Kolenda et al. (325), Alonso-Caballero et al. (326) and Knight et al. (183).

Many attempts haven been made over the last several decades by several researchers, such as Hurst et al (327), to find a site of colonization in grass grubs, but all have been unsuccessful, therefore it could be possible that Sef does not play a role in adherence to cells in the grass grub gut, but instead play a role in binding to the material that grass grub ingest or plant matter around which the grass grubs reside. It is also plausible that these fimbriae serve a non- insect related function.

Table 6-3: Closest unique functionally described orthologs for SefH-SefJ obtained using BlastX. Hypothetical proteins and redundant proteins were omitted.

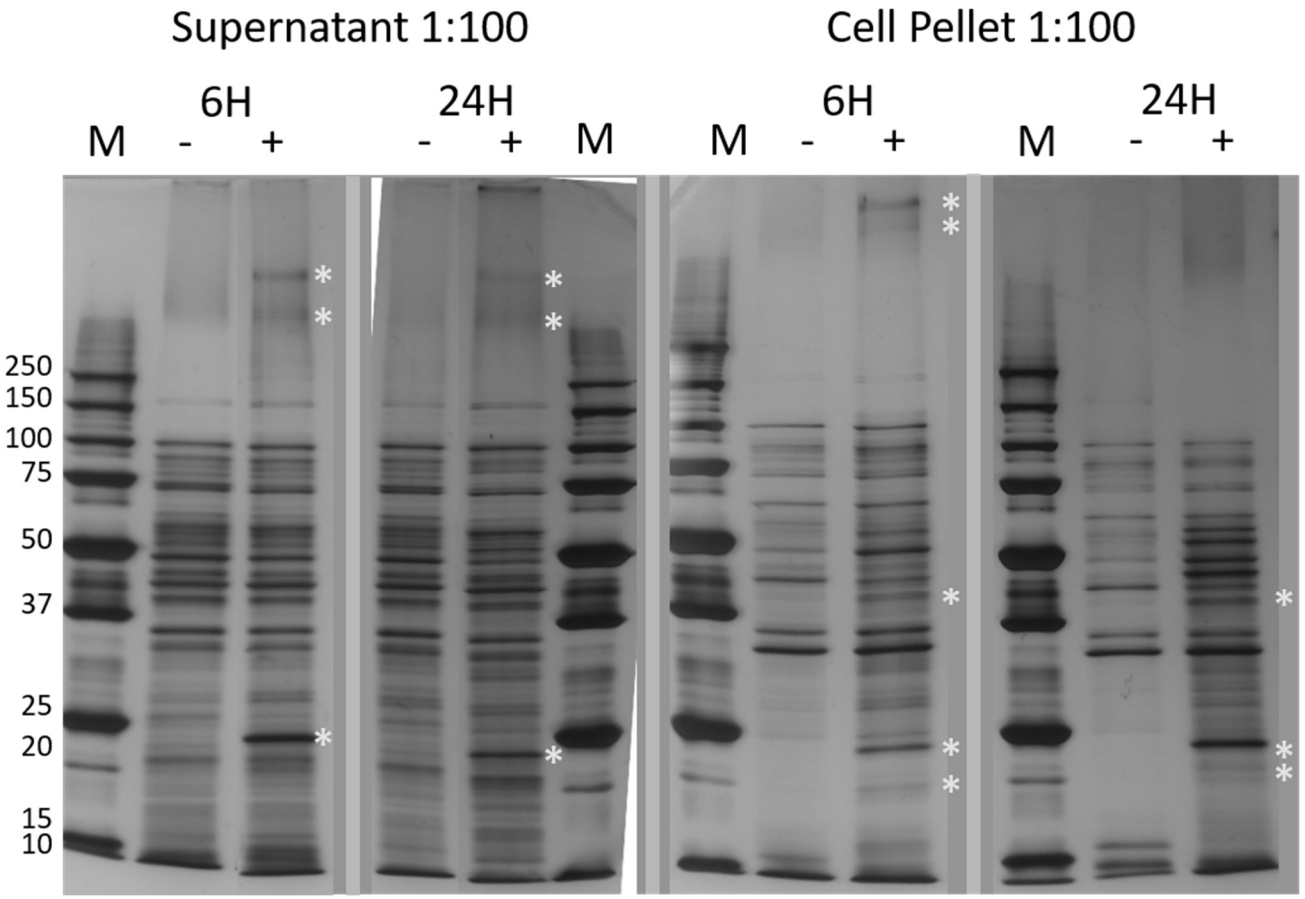
Protein	Hit						Accession	
SefH	MULTISPECIES: fimbrial protein [*Providencia*]						WP_109913237.1	
174 AA residues	Length:	170	Range:	25 to 170	Eval:	6E-55		
	Ident:	85/146(58%)	Pos:	106/146(72%)	Gap:	0/146(0%)	Coverage:	83.0%
	fimbrial minor subunit StfF [*Salmonella enterica* subsp. Salamae]						ECG8606686.1	
	Length:	158	Range:	7 to 158	Eval:	1E-33		
	Ident:	71/159(45%)	Pos:	99/159(62%)	Gap:	9/159(5%)	Coverage:	90.0%
SefI	adhesin [*Morganella morganii*]						WP_107679199.1	
270 AA residues	Length:	272	Range:	158 to 272	Eval:	4E-09		
	Ident:	44/117(38%)	Pos:	61/117(52%)	Gap:	6/117(5%)	Coverage:	41.0%
	adhesin [*Proteus columbae*]						WP_100158730.1	
	Length:	275	Range:	162 to 275	Eval:	6.00E-09		
	Ident:	46/117(39%)	Pos:	64/117(54%)	Gap:	8/117(6%)	Coverage:	41%
SefJ	adhesin [*Proteus columbae*]						WP_100158730.1	
197 AA residues	Length:	275	Range:	153 to 273	Eval:	4.00E-05		
	Ident:	39/124(31%)	Pos:	66/124(53%)	Gap:	6/124(4%)	Coverage:	61.0%
	adhesin [*Proteus mirabilis*]						WP_088493555.1	
	Length:	275	Range:	153 to 273	Eval:	2.00E-04		
	Ident:	39/123(32%)	Pos:	65/123(52%)	Gap:	4/123(3%)	Coverage:	61%



6.3 **Fimbriae expression**

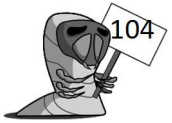
To determine if the Sef fimbriae were functional, the region from *sefA* to *sefJ* was cloned into an arabinose induction vector, labelled inhouse as pAY2-4 to form pARA_Sef (Section 2.2.3, Supplementary Materials Appendix AS.8). The vector was transformed into the Δfim *E. coli* strain AAEC072A, also known as MG1655ΔfimA-H (217). This strain does not encode for a Type 1 fimbria. To confirm Sef was being expressed, AAEC072A containing pAY2-4, and AAEC072A containing the Sef expressing pARA_Sef vector, were both grown for 6 h and 24 h under arabinose induction, as outlined in Section 2.2.20. At these times, the cell pellet and supernatant of culture were assessed for the presence of Sef components using Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE), as outlined in Section 2.2.21. Visual assessment of the resultant SDS-PAGE gel revealed the presence of several bands that were of similar sizes to the predicated sizes of several Sef proteins present in the AAEC072A pARA_Sef that were absent in the AAEC072A pAY2-4 induced cultures (Figure 6-5).

Assessment of the induced cultures, grown o/n for approximately 24h, under Transmission electron microscopy (TEM) revealed that the Sef expressing *E. coli* strain AAEC072 (pARA_Sef), indeed produces fimbriae, as well as flagellum structures, whereas *E.coli* strain AAEC072, containing only the pAY2-4 vector, only produces the flagellum (Figure 6-6). The flagellum is a filamentous apparatus, used to provide locomotion (328) and sometimes sensory properties to the bacteria (329). Based on these observations, the fimbriae expressed in AAEC072A with pARA_SEF are thus the Sef derived from the Sef encoding operon. Additional comparative TEM images can be found in Supplementary Materials S.22



Protein	kDa
SefA	17.58
SefB	19.2
SefC	92.81
SefD	27.35
SefE	22.85
SefF	18.52
SefG	29.37
SefH	18.64
SefI	28.16
SefJ	20.46

Figure 6-5: SDS-PAGE cultured supernatant and cell pellet comparison between ΔfimA-H *E. coli* strain AAEC072A (217) containing empty arabinose induction vector pAY2-4 (-), and said Δfim *E. coli* transformed with the Sef expression vector pARA_SEF (+). Cultures were induced with 0.2 mg/ml arabinose. Samples were taken at 6 h and 24 h. A 1:100 dilution of both cell pellet and supernatant of each timepoint for both WT and Sef expressing mutant was put on an SDS-PAGE gel. Arrows denote location of possible fimbrial genes, refer to protein table (right) for predicted sizes. M = marker, band sizes of marker are indicated on left hand side.



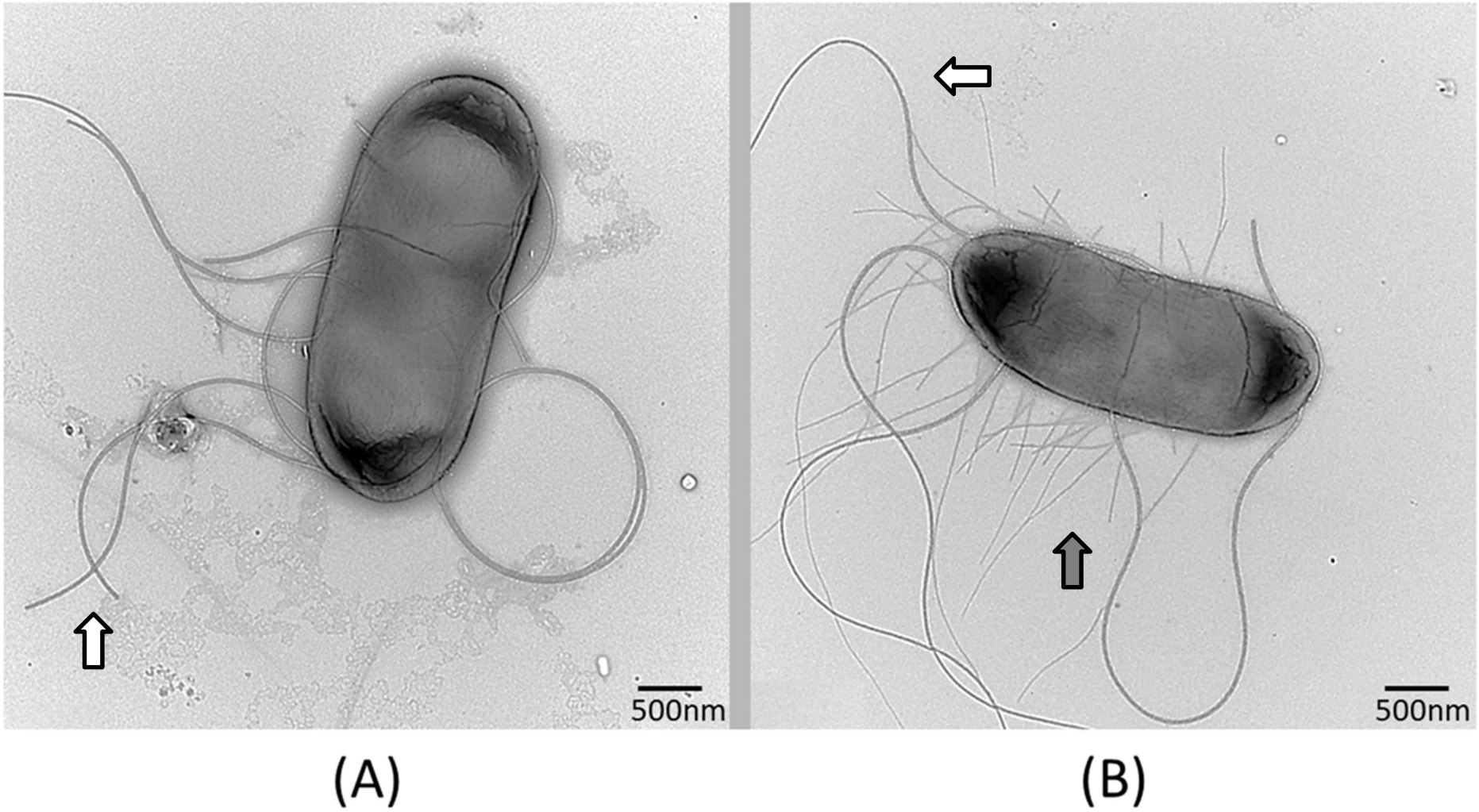


Figure 6-6: Electron micrograph of the ΔfimA-H *E. coli* strain AAEC072A (pAY2-4) (217), and AAEC072 (pARA_SEF). A) AAE072CE (pAY2-4), a fimbriae null *E. coli* strain. B) AAE072CE (pARA_SEF), containing the whole SEF operon and a Cm cassette for screening purposes. Grey arrow denotes Sef, White arrow denotes flagellum. Refer to Supplementary Materials S.22 for additional examples.

6.4 **Sef contribution to pathogenicity**

Due to the correlation between Sep and Sef as can be observed in Figure 6-3, it is likely Sef is involved with virulence in some unknown way. To determine the effect of Sef on the pathogenicity of pADAP-bearing *Serratia* species against grass grub, an A1MO2 *sefA-C* deletion variant was constructed as outlined in Section 2.2.11. For assessment of fimbrial production, the isolates were grown o/n in LB-broth and assessed the following day by TEM. The A1MO2-isolate containing the *ΔsefA-C* pADAP mutant showed no fimbriae formation under TEM (Figure 6-7). Approximately 100 cells were assessed per sample. No pADAP*ΔsefA-C* cells were found to have fimbria. A similar assessment of A1MO2 WT found, that approximately 10% of bacteria had fimbria. Based on these observations, it can be concluded that the fimbrial deficient mutant does not produce any fimbria, further validating the *sef* cluster as encoding the observed fimbriae. The observed presence of fimbriae in only 10% of the WT population suggests that there could be an additional trigger needed for proper expression of the Sef fimbriae structure.

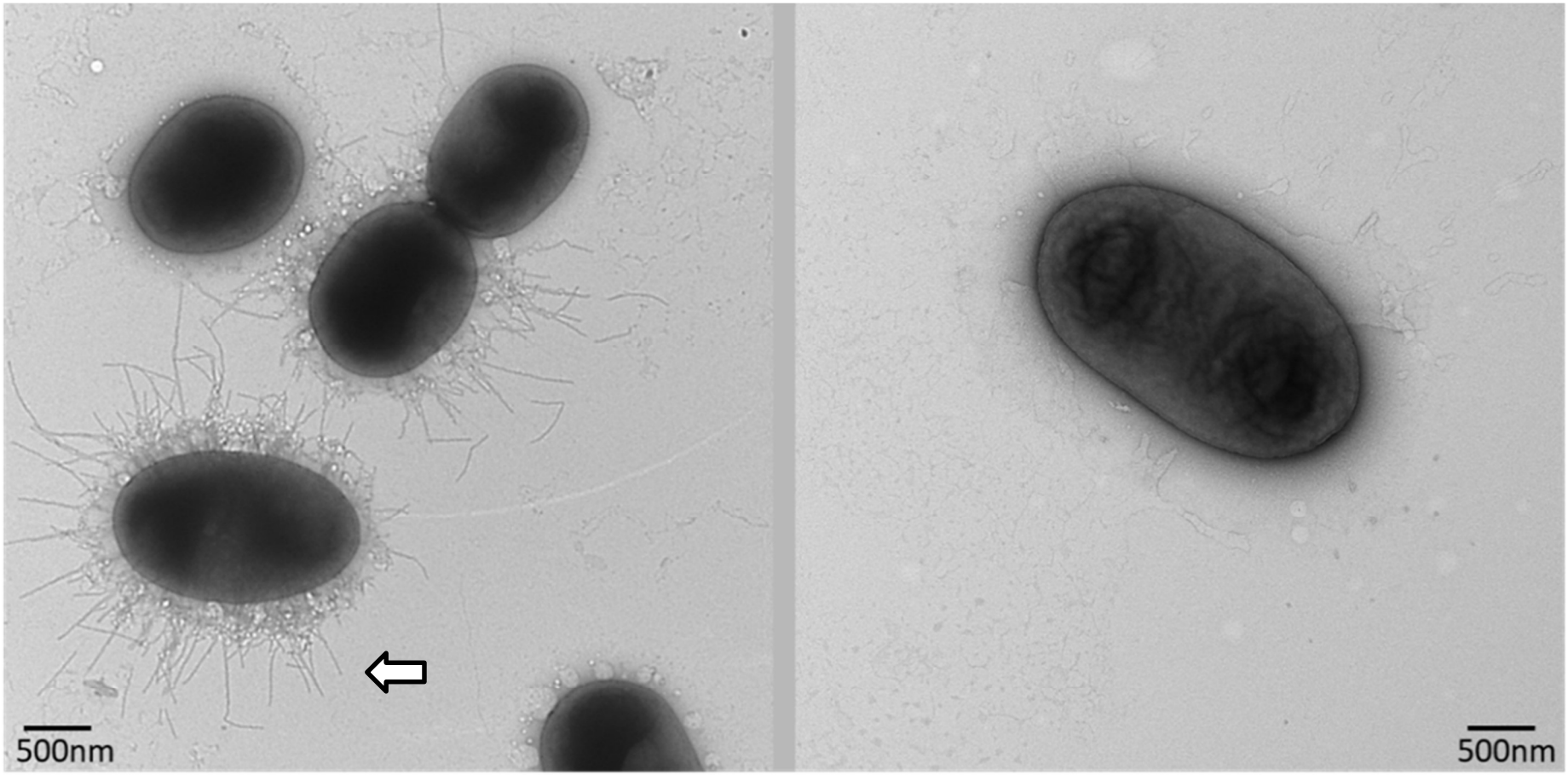


Figure 6-7: Comparison between A1MO2 WT (pADAP) and its A1MO2 pADAPΔ*sefA-C* mutant derivative. A) Electron micrograph of A1MO2 with pADAP. B) Electron micrograph of pADAPΔ*sefA-C* mutant in which *sefA, B* and *C* were deleted. White arrow denotes Sef.

6.4.1 Bioassay results Sef null A1MO2 mutant

To determine if the A1MO2 Δ*sefA-C* mutant affected virulence, a dose response of A1MO2 WT and the A1MO2 with pADAP*ΔsefA-C* cultures were independently assessed by bioassay. Bioassays were performed in duplicate. Assessment of resultant bioassay data showed that there was no difference between the Sef producing WT and the Sef null mutant, with the predicted LD_{50} of 10^7 cells per larvae for each strain. Based on this, it can be concluded that within the confines of the bioassay setup, the Sef has no direct host associated effect (Table 6-4).

Table 6-4: Comparative bioassays of A1MO2 WT and the pADAPΔsefA-C mutant carrying A1MO2 isolate.

Strain	Disease%	Mortality%	Affected%
Blank	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
pADAP 10^9	91.7 ± 5.8	8.3 ± 5.8	100.0 ± 0.0
pADAP 10^8	66.7 ± 9.8	4.2 ± 4.2	70.8 ± 9.5
pADAP 10^7	50.0 ± 10.4	0.0 ± 0.0	50.0 ± 10.4
pADAP 10^6	4.2 ± 4.2	4.2 ± 4.2	8.3 ± 5.8
pADAP 10^5	4.2 ± 4.2	4.2 ± 4.2	8.3 ± 5.8
pADAPΔsefA-C 10^9	87.5 ± 6.9	12.5 ± 6.9	100.0 ± 0.0
pADAPΔsefA-C 10^8	75.0 ± 9.0	12.5 ± 6.9	87.5 ± 6.9
pADAPΔsefA-C 10^7	45.8 ± 10.4	0.0 ± 0.0	45.8 ± 10.4
pADAPΔsefA-C 10^6	4.2 ± 4.2	0.0 ± 0.0	4.2 ± 4.2
pADAPΔsefA-C 10^5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Based on the lack of observed difference in dose response between pADAP and pADAPΔsefA-C encoding A1MO2 strains, it is plausible that the Sef has an alternate function, such as binding to a nutritional source that allows the *Serratia* to more easily be transported from the soil into the gut of the larvae. Alternately, the Sef acts as an anchor to a location frequented by grubs in the rhizosphere or it could help with adherence to other Serratia cells to increase colonization potential. To determine if the carrot used in bioassays is a potential binding target, A1MO2 bearing a GFP producing vector in conjunction with pADAP or pADAPΔsefA-C were assessed for adherence through fluorescent microscopy.

6.4.2 Adherence to carrot

To determine if the adherence is associated with a food source, adherence of bacteria to carrot slices was compared between A1MO2 WT (pADAP) and the A1MO2 with pADAP*ΔsefA-C*, using the method as outlined in Section 2.2.22. These strains were transformed with an additional pUC30TGFPMut3 vector that encodes for a GFP. Images were taken with the microscope, as outlined in Section 2.2.23, before (Figure 6-8) and after (Figure 6-9) washing with MilliQ ddH2O for 1 h. With reference to these figures, it is evident that both the pADAP and the pADAP*ΔsefA-C* mutant carrying cells have similar levels of cells per surface before and after washing. High numbers of fluorescent cells were observed in the liquid after washing in both strains with accordingly fewer fluorescent cells observed on the carrot slices after washing. The diminished fluorescence is therefore an effect of detachment from the carrot.

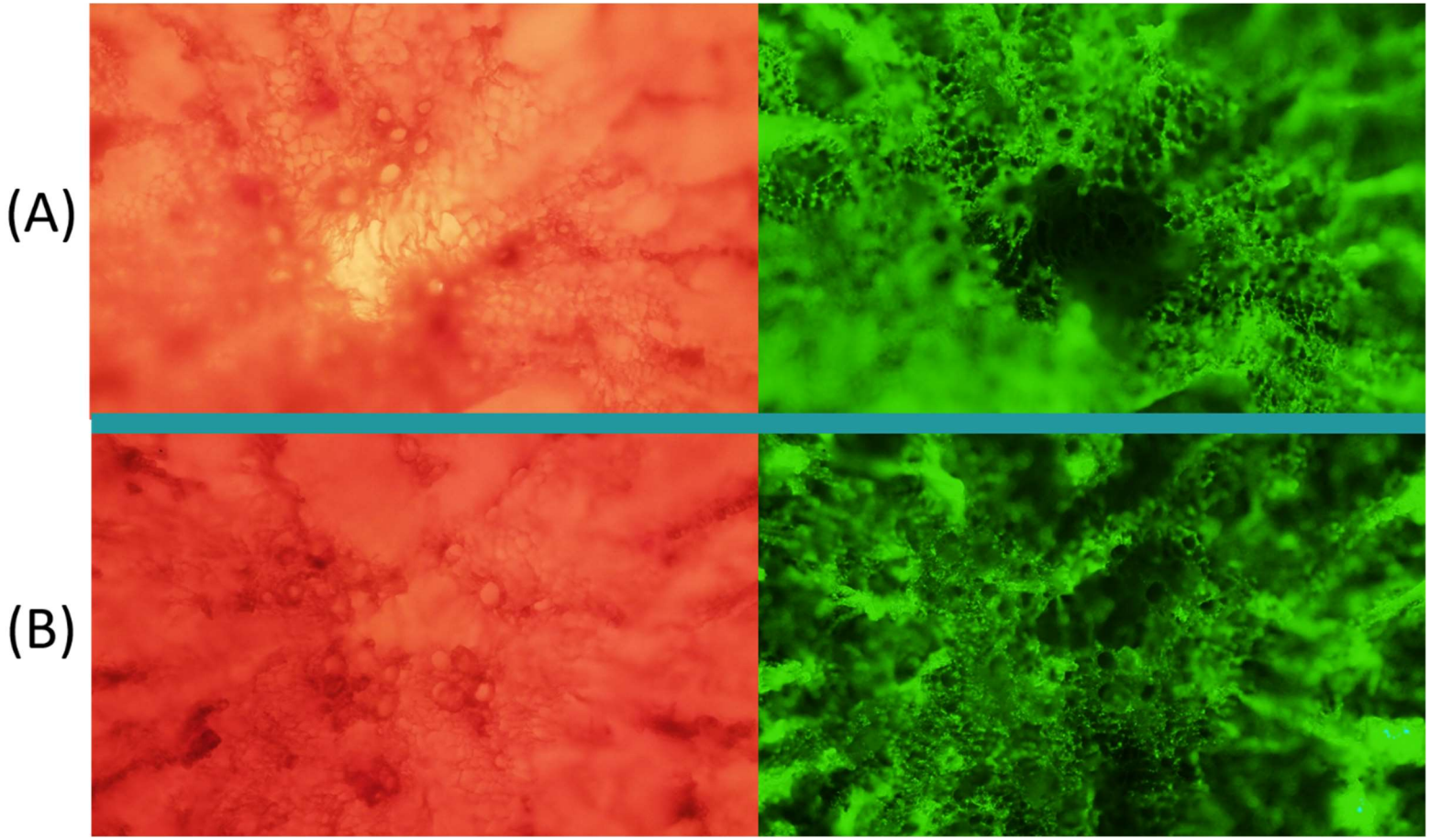


Figure 6-8: 100x microscope image of the Xylem after ~16 h of incubation. A) Carrot slice inoculated with 5 µl of A1MO2 WT. B) Carrot slice inoculated with 5 µl of A1MO2 carrying pADAPΔ*sefA-C*

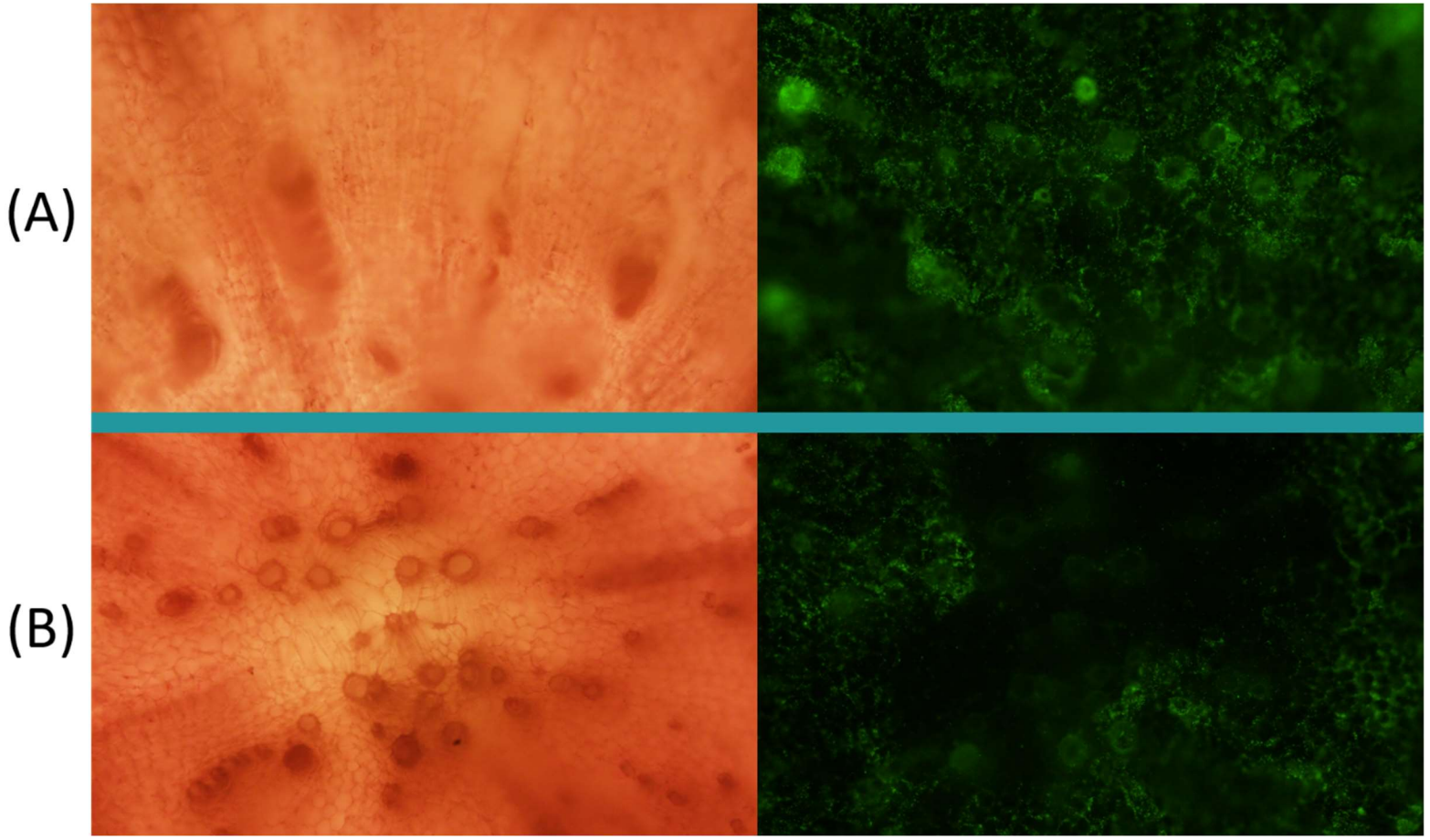


Figure 6-9: 100x microscope image of the Xylem after 1 h of washing in MilliQ ddH2O. A) Carrot slice inoculated with 5 µl of A1MO2 WT. B) Carrot slice inoculated with 5 µl of A1MO2 carrying pADAPΔ*sefA-C*

6.4.3 Oral injection of Sef defficient bacteria

To determine if fimbriae were involved in adherence to food post ingestion, a 1:10 dilution of the A1MO2 WT (pADAP), and A1MO2 (pADAP*ΔsefA-C*) samples, ~10^8 cells each, was independently injected into the oesophagus of 3rd instar grass grub larvae as described in Section 2.2.23. A subset of larvae was left unfed, and the others were provided with a ~3 mm^3 carrot cube post injection, to see if there was any adherence to carrot inside the gut, as an alternate hypothesis for colonization location and binding affinity of Sef. The following day the guts were dissected out of each larva (Figure 6-10) and assessed under fluorescent microscope.

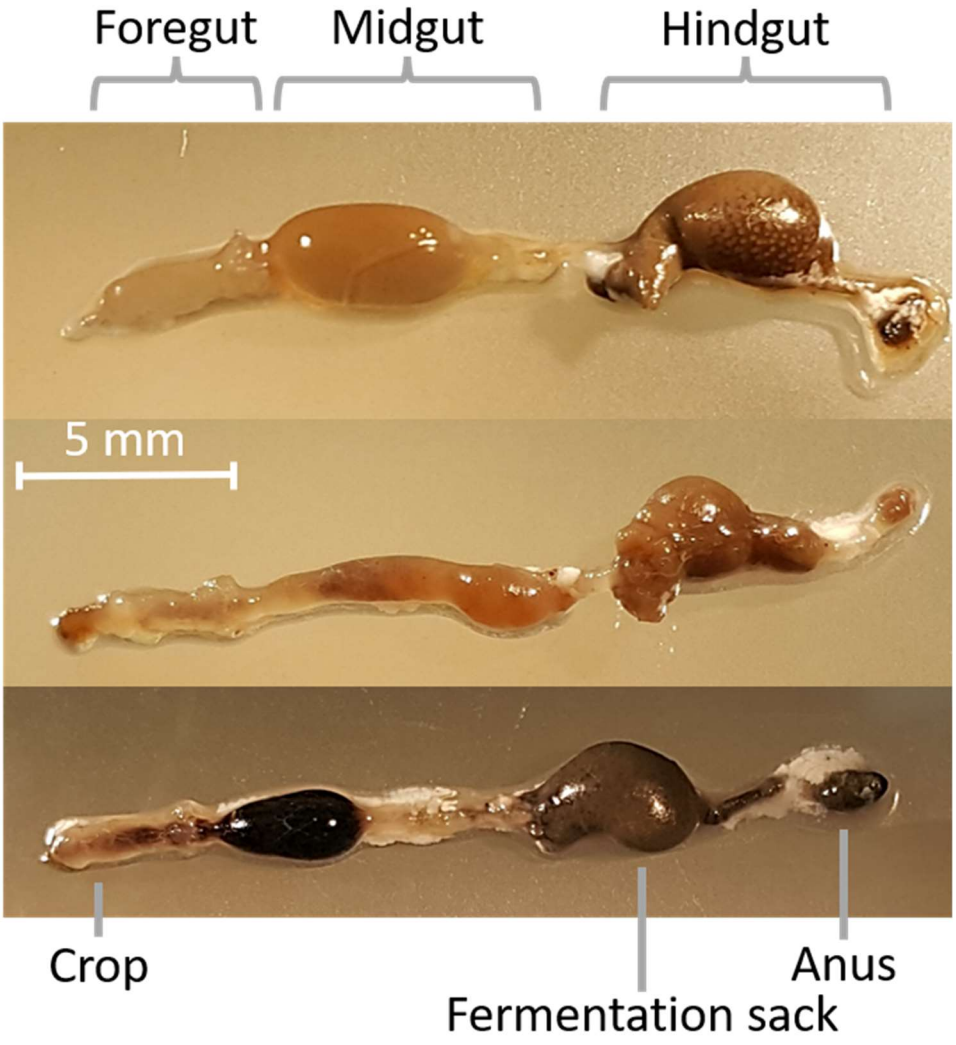


Figure 6-10: Several examples of extracted grass grub guts used to assess for Sef binding. Areas of the intestine are indicated as described by Jackson et al. (330).

For microscopy the guts were laterally dissected for proper view of the inside milieu. Images were taken from several regions in the foregut/midgut but particularly the hindgut where food material is transported to for processing (Figure 6-11).

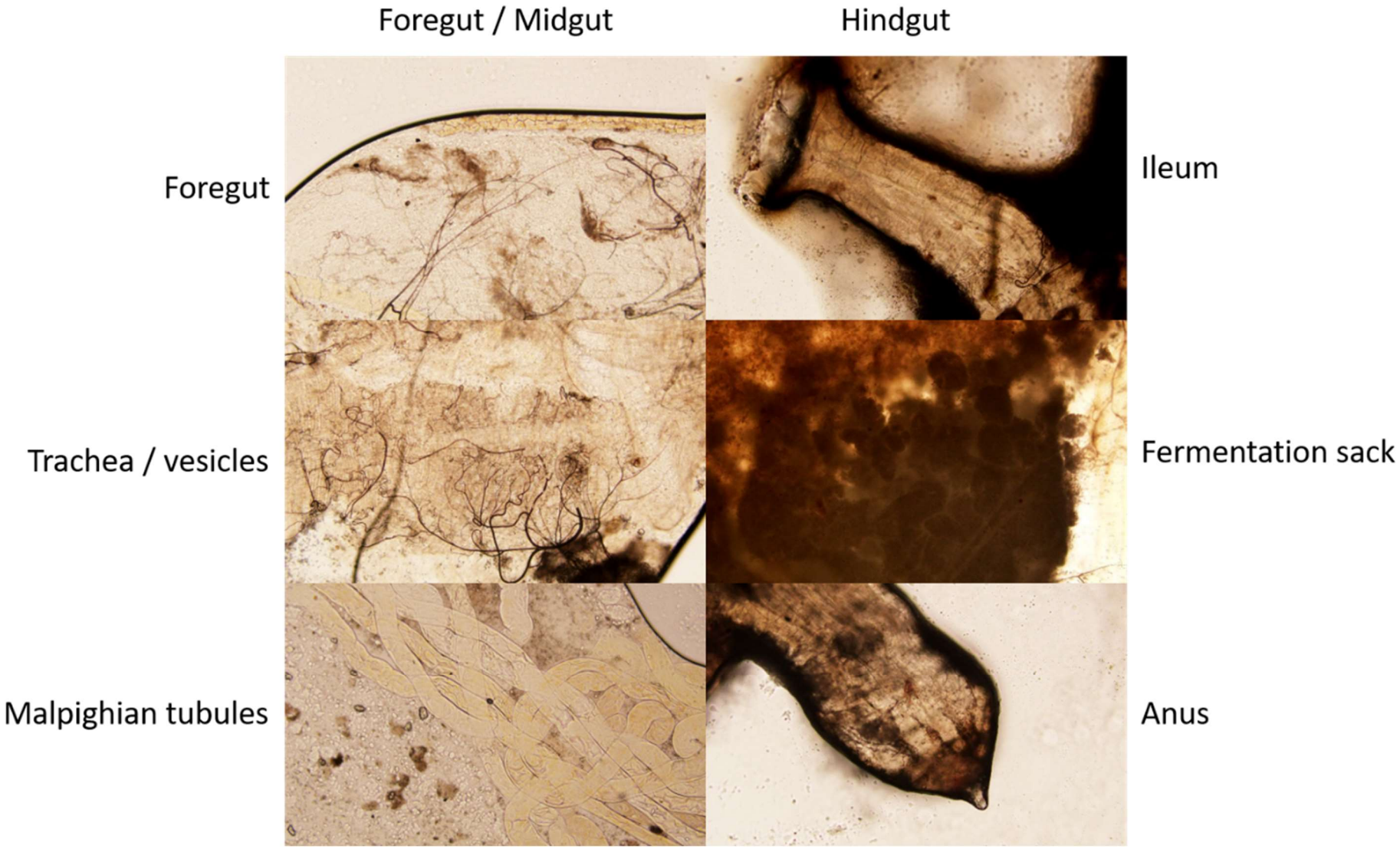


Figure 6-11: Examples of the regions analysed for localization of fluorescent cells. Images were taken at 100x magnification (10x objective, 10x lens).

Unfortunately, very few fluorescent cells were found, similar to previous published data by Hurst et al. (327). No localized fluorescence was observed inside the gut, with most fluorescent cells being found in the liquid contents of the larval gut. In addition to this, no adherence to food particulates was observed, for either the pADAP bearing or the pADAP*ΔsefA-C* bearing strain. The only observed clustering of fluorescent cells occurred inside the Malpighian tubules (Figure 6-12), however this is a region known for autofluorescence (331) and therefore no distinction can be made between bacterial cells or auto fluorescent events. Unfortunately, these pUC30TGFPMut3 bearing strains were only constructed in the final stages of the PhD at the end of the *C. giveni* season. As previously mentioned in Section 1.1, *C. giveni* is holometabolic, its larvae stage is only found in the months approximately between January and June, unfortunately at the time of this experiment it was June, meaning at this time, there were no healthy grass grubs to undertake additional imaging on.

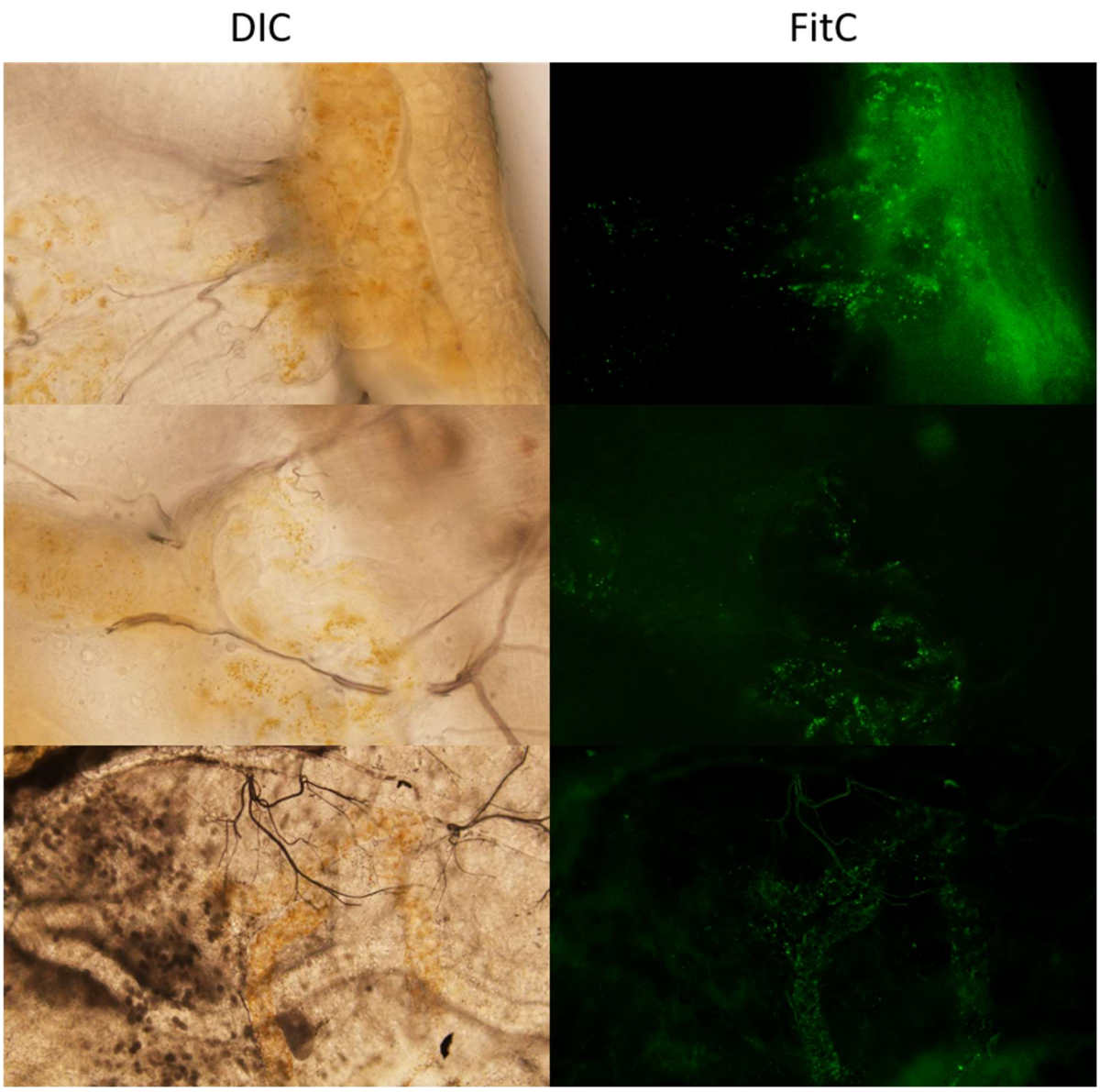


Figure 6-12: Fluorescent cells in and around the Malpighian tubules.

Chapter 7 Antifeeding prophage

As outlined in Section 1.12, the Afp region found on the pADAP plasmid encodes a R-type tailocin-like apparatus (161, 204), that is morphologically similar to type VI secretion systems (206), R-type pyocins [184] and *Photorhabdus* virulence cassettes (PVC) structures (191). These apparati comprise a base-plate, outer sheath and an inner tube-like structure with a pointed ending, and tail fibres (194). Through TEM, the Afp was observed to exist in two forms, extended and contracted. Through contraction of the outer sheath, the particle can forcefully protrude the inner tube (161) and potentially inject virulence determinants. A schematic of the Afp structure and a model of how it operates is depicted in Figure 1-10.

7.1 Afp PAI

The Afp PAI consists of two regulatory genes, 16 open reading frames (ORF) transcribing the proteins needed for the Afp construction, and 2 ORFs encoding the virulence determinants (198). The translated products of Afp ORFs *afp1-16* share a high AA similarity to several PVC found in *Photorhabdus* bacteria associated with pathogenic species of insect active nematodes (191). A full list of ORFs encoding the Afp and their function in the formation of the Afp particle can be found in Supplementary Materials S.4. The Afp causes cessation of feeding in the *C. giveni* larvae (161). Two Afp variants were previously published, the second being named AfpX which is known to cause rapid death in the grass grub, termed hyper-pathogenicity (69). In this study a third Afp type particle has been identified: it resides on the pADAP-type plasmid found in *S. proteamaculans* isolate 1137, that was isolated from samples obtained from around Ashburton, Canterbury (NZ). This novel Afp is designated as AfpA.

This *S. proteamaculans* 1137 isolate is of interest as, based on sequence similarity of the backbone, the plasmid is one of the most diverged from all others (Figure 4-7). Although there is substantial nucleotide as well as amino acid dissimilarity between most ORFs, aside from *afp17*, the entire Afp region appears to be present. The most divergent ORFs are *afp13A* and *afp18A*, with 29.7% and 47,2% amino acid similarity to their Afp counterparts respectively, eluding to different host range and toxin. The AA alignments of Afp13A to Afp13 and that of Afp18A to Afp18 are shown in Supplementary Materials S.26. A schematic of the three Afp variants is presented in Figure 7-1, and the nucleotide distance matrix between all of the Afp regions is depicted in Figure 7-2.

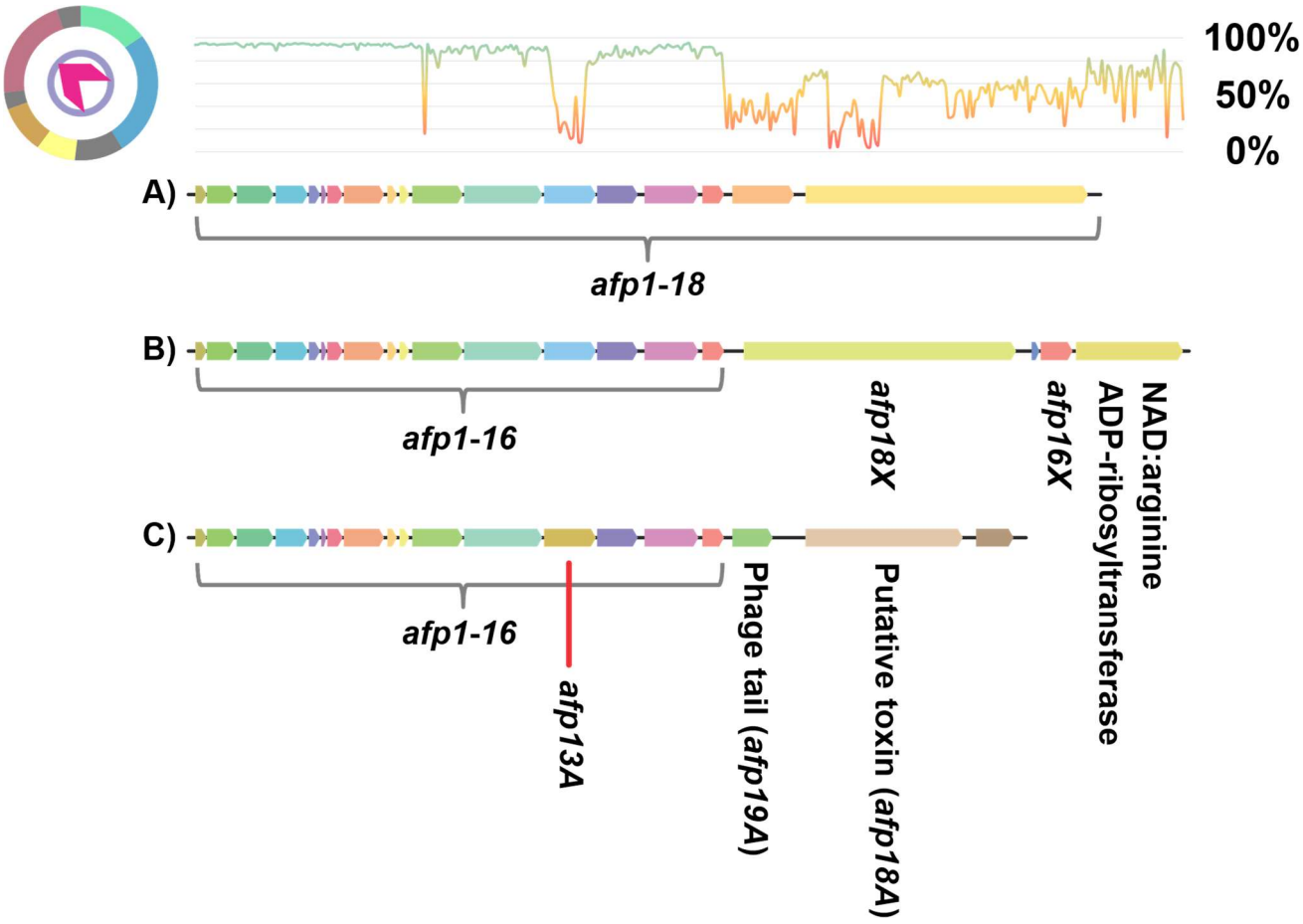


Figure 7-1: The three Afp encoding regions currently known. A and B have been previously published as Afp (198) and AfpX (69), the region of C, identified in this study, is only present in one isolate, 1137. The C variant, designated as AfpA, shows remarkable nucleotide and amino acid sequence dissimilarity compared to both other PAIs (<70%). AfpA lacks an *afp13*, *afp17* and *afp18* homolog.

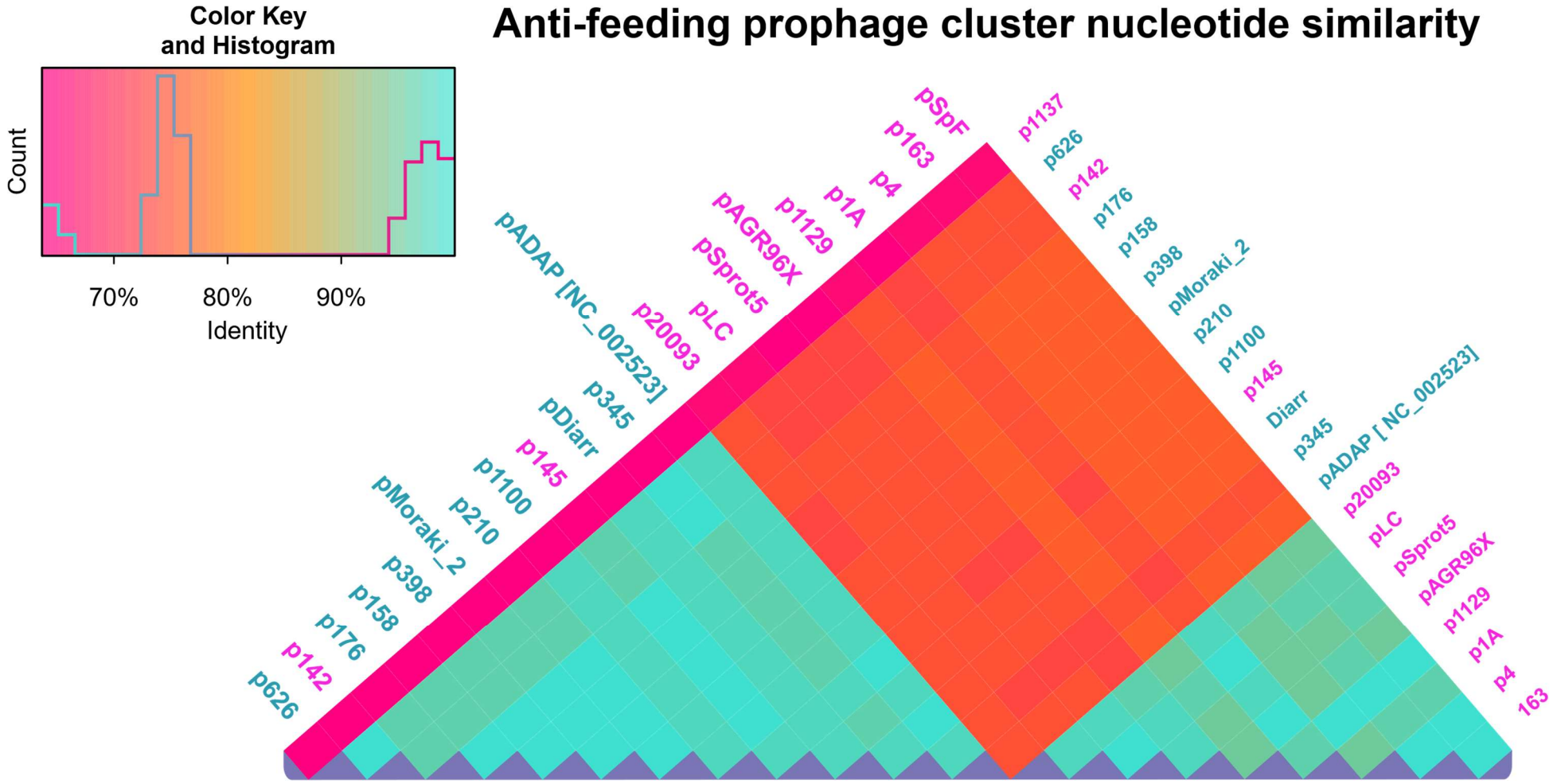


Figure 7-2: Distance matrix of all Afp PAIs analysed in this study. There is a clear separation of nucleotide conservation between the Afp, AfpX and the novel Afp found in this study. The *S. proteamaculans* 1137 AfpA is the most dissimilar region and does not cluster with either Afp variant. Blue labelled strain names are *S. entomophila* and pink *S. proteamaculans*.

7.2 Afp distribution

Twenty-two pADAP-type plasmids were found to encode the Afp PAI. Ten of these came from *S. entomophila* and the remainder from *S. proteamaculans*. The distribution of these Afp PAIs per species, based on the isolates assessed in this study, is listed in Table 7-1.

Table 7-1: Distribution of isolates encoding pADAP Afp variants. The C type found in isolate 1137 has been named AfpA.

Type[1]	Genus	Species	Isolate
A (Afp)	*Serratia*	*entomophila*	158, 176, 210, 345, 398, 626, 1100, Diarr, Moraki_2, pADAP
		proteamaculans	142, 145
B (AfpX)	*Serratia*	*proteamaculans*	4, 163, 1129, 1A, 20093, SpF, LC, AGR96X, Sprot5
C (AfpA)	*Serratia*	*proteamaculans*	1137

[1] Type from Figure 7-1

Table 7-1 reveals some surprising results, firstly, the Type (A) Afp PAI, apart from the *S. proteamaculans* isolates 142 and 145, is exclusively found in the chronic disease inducing *S. entomophila* isolates. As previously outlined in Section 4.9, the plasmids obtained from *S. proteamaculans* isolates 142 and 145 are highly homologous to *S. entomophila* pADAP plasmids in both the backbone, as seen in the DNA similarity matrix (Figure 4-7), as shown in the DNA similarity matrix of the Sep encoding region (Figure 5-1) and likewise with the presence of the *S. entomophila* Afp (Figure 7-2). Chapter 12 contains more information on the pADAP plasmids found in these two *S. proteamaculans* isolates. Another more interesting result however is that AfpX particle, that previously was assumed to be exclusively causal for hyper-pathogenicity in grass grub larvae, appears to be present in several isolates that do not appear as hyper-pathogenic as the AGR96X type strain. Apart from isolate 1129, the plasmids carrying the AfpX PAI are identical to each other both in backbone (Figure 4-7), as well as all plasmids sharing the (G) type demarcation region (Figure 4-15). Except for p1129, none of these plasmids encode for Sep. Therefore, it is surprising that not all these plasmid bearing isolates, most notably 4, 163 and Sprot5, produce the same bioactivity phenotype, suggestive of a chromosomal component to the bioactivity.

Table 7-2: Bioactivity of AfpX carrying isolates on grass grub. * Strain 1129 is the only strain that encodes the Sep PAI and deviates from the rest by having the type F demarcation region instead of G like the rest.

Isolate	Disease%			Mortality%			Affected%		
4	92	±	8	8	±	8	100	±	0
163	17	±	8	33	±	10	50	±	10
1129*	21	±	8	75	±	9	96	±	4
1A	0	±	0	80	±	13	80	±	13
20093	5	±	8	90	±	8	95	±	8
AGR96X	10	±	8	90	±	8	100	±	8
LC	11	±	4	88	±	5	98	±	2
SpF	35	±	8	60	±	8	95	±	8
Sprot5	50	±	8	40	±	8	90	±	8

Alternately it is also plausible that isolates such as 4, 163 and Sprot5, with lower mortality rates, maybe regulated differently through some yet to be determined function. Further study into the chromosomal aspects of these strains is currently being undertaken by Amy Vaughan [unpublished data], to understand the full scope of the system.

7.3 Novel Afp particle found in isolate 1137

To determine if an Afp particle is produced by the 1137 isolate, the AfpA was induced and purified as outlined in Section 2.2.18. After purification, the samples were analysed under TEM to observe the morphology of the particles. Many particles were found to be in contracted state, and the few that were uncontracted seemed to have a slightly asymmetrical shape. The particles were found to be ~ 100 nm, similar to Afp and AfpX. The AfpA particles can be seen in Figure 7-3.

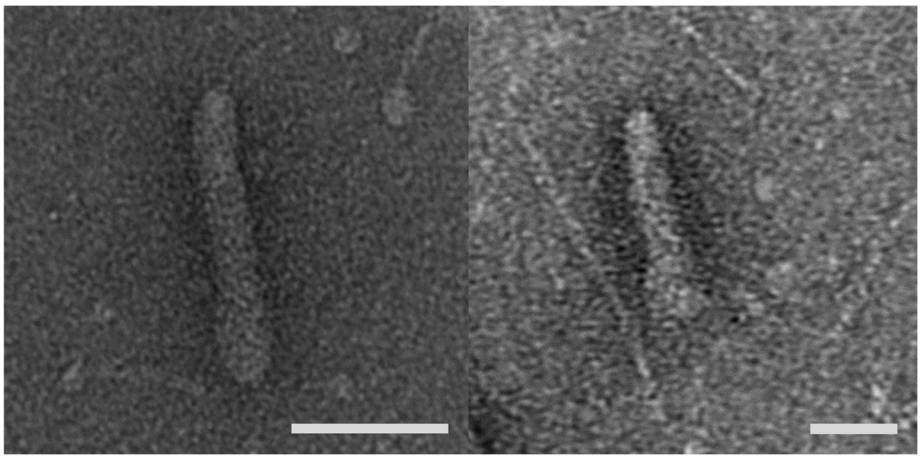


Figure 7-3: Two electron micrographs of the novel AfpA particle extracted from isolate 1137. Scale bars represent 50 nm. The particles are not as symmetrical as the Afp and AfpX particles, and the tail fibres are not visible although these might have been damaged in the purification process.

Post ingestion by grass grub larvae, the Afp particle causes cessation of feeding activity, and is one of the symptoms of amber disease that can persist up to several months. Amber disease eventually leads to the death of the larvae due to a combination of malnourishment and sepsis, followed by bacterial invasion of the hemocoel (81, 208). It has been shown that purified particles of Afp can cause the antifeeding phenotype, and that ~500 particles are required for an LD_{50} (201). Bioassay assessments of *S. proteamaculans* isolate 1137 revealed no bioactivity in grass grub (Table 7-3). This likely reflects a difference in the tail fibres or toxins encoded by this Afp variant.

Table 7-3: Percent disease and mortality of *S. proteamaculans* 1137 compared to *S. entomophila* A1MO2 (pADAP) and *S. proteamaculans* AGR96X (pAGR96X) and untreated control towards grass grub larvae.

Strain	Disease%	Mortality%	Affected%
Control	2.4 ± 1.7	1.2 ± 1.2	3.6 ± 2.0
A1M02	83.3 ± 4.1	3.6 ± 2.0	86.9 ± 3.7
AGR96X	10.0 ± 9.0	90.0 ± 9.0	100.0 ± 0.0
1137	0.0 ± 0.0	2.8 ± 2.8	2.8 ± 2.8

Based on the observed difference of the tail fibres and toxins of AfpA, the *S. proteamaculans* 1137 isolate was also bioassayed against other members of the Scarabaeidae family to find a possible alternative host. These included Red headed cockchafer (*Adoryphorus couloni*), Manuka beetle (*Pyronata festiva*) and Tasmanian grass grub (*Acrossidius tasmaniae*). A picture of these larvae can be seen in Supplementary Materials S.9. These bioassays showed no bioactivity in any of the larvae exposed to the 1137 isolate (Table 7-4).

Table 7-4: Percent disease and mortality of the *S. proteamaculans* 1137 isolate in non-grass grub insects. Red headed cockchafer (*Adoryphorus couloni*), Manuka beetle (*Pyronata festiva*), Tasmanian grass grub (*Acrossidius tasmaniae*). No apparent response can be observed.

Bioactivity	Disease%	Mortality%	Affected%
Red headed cockchafer	0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Manuka beetle	0 ± 0	10 ± 10	10 ± 10
Tasmanian grass grub	0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Based on bioassay results, it is plausible that the particle does not have the appropriate tail fibres to adhere to grass grub or the other tested scarab species. It is also plausible that the p1137 encoded Afp18A toxin is non-active or does not affect grass grub, as the Afp18A is only 47.2% similar to Afp18 on an amino acid level. The putative toxin in the p1137 encoded AfpA has no other functionally annotated orthologs and is only ~25% homologous to AfpX18 (Table 7-5). The translated sequence of the two separate tail fibre ORFs also have no direct ortholog in the RefSeq database (Table 7-6). Based on these findings, it is likely that the AfpA particle has a yet to be defined target. Full alignment of Afp13A to Afp13 and Afp18A to Afp18X can be found in Supplementary Materials S.26.

An alternate explanation to absence of bioactivity to the tested scarab species is that the production of this novel AfpA is halted by the absence of a functional regulatory operon that is further discussed in the next section.

Table 7-5: Annotation of the predicted AfpA toxin based on BlastX hits.

Protein	Hit				Accession	
Afp18A	MULTISPECIES: hypothetical protein [*Pectobacterium*]				WP_039484771.1	
1949 AA residues		2188	7 to 1399	0		
		832/1431(58%)	1012/1431(70%)	46/1431(3%)	58.1%	
	AfpX18 [*Serratia proteamaculans*]				ANV21616.1	
		2285	810 to 1691	0		
		415/934(44%)	562/934(60%)	91/934(9%)	44.4%	

Table 7-6: Annotation of Afp13A and Afp19A based on BlastX hits.

Protein	Hit						Accession	
Afp13A	Afp13 [*Serratia entomophila*]						AAT48350.2	
512 AA residues	Length:	636	Range:	1 to 437	Eval:	8E-55		
	Ident:	172/470(37%)	Pos:	241/470(51%)	Gap:	49/470(10%)	Coverage:	36.6%
	tail fiber protein [*Serratia fonticola*]						WP_021808099.1	
	Length:	396	Range:	79 to 274	Eval:	4E-45		
	Ident:	104/210(50%)	Pos:	133/210(63%)	Gap:	23/210(10%)	Coverage:	49.5%
Afp19A	phage tail sheath family protein [*Chromobacterium haemolyticum*]						WP_081548074.1	
555 AA residues	Length:	564	Range:	224 to 562	Eval:	7E-93		
	Ident:	161/348(46%)	Pos:	220/348(63%)	Gap:	17/348(4%)	Coverage:	46.3%
	phage tail sheath family protein [*Photorhabdus* sp. CRCIA-P01]						WP_118985780.1	
	Length:	486	Range:	87 to 483	Eval:	2E-92		
	Ident:	170/412(41%)	Pos:	246/412(59%)	Gap:	20/412(4%)	Coverage:	41.3%

7.4 Afp Regulatory genes

Both the Afp and AfpX bearing pADAP-type plasmids have an upstream regulatory region that is considered to regulate transcription and release of the Afp particles (Figure 7-4). This entire region is absent from p1137 and could potentially be the reason behind the absence of bioactivity with this isolate.

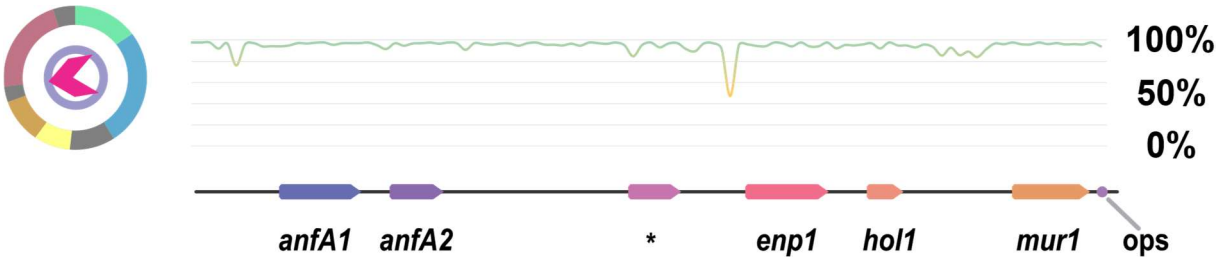


Figure 7-4: The Afp and AfpX regulatory region, this region is entirely absent from p1137 with the exception of the 11bp ops sequence. This region contains two holin genes, and two lysis genes as well as an anfA1 gene that is required for transcription of the full Afp operon. * denotes ORF with no known hit in RefSeq database.

The *anfA1* ORF has been shown to be linked to the production of AFP particles (332). The *anfA1* ORF is a transcription elongation factor, which is required for the proper transcription of long regions (333) by preventing backtracking and hairpin-stabilized pausing of RNAPs (311). The Afp operon is ~27 Kb in size and therefore it is understandable that without a mechanism to stabilize transcription of the entire region, transcriptional factors will stop early on. Therefore, the *anfA1* shares a similar function to the *sea3*/*nusG* factor (outlined in section 4.12), in that they facilitate transcription of long regions, and this is reflected in low level homology to NusG (Table 7-7). Besides *anfA1* there are several genes related to lysis and secretion such as *anfA2*. The AnfA2 encoding ORF appears to encode for a 95 AA protein, with similarity to a Lambda holin family, part of the Holin Superfamily III (334). Holins are required to transport endolysins Enp1 from the cytoplasm into the periplasm in gram-negative bacteria that have both an outer and inner membrane (335). These endolysins then in-turn destabilized the peptidoglycan in the periplasm (336). Once peptidoglycan disruption has occurred, lysis of the cell can happen through the muramidase-type lysozyme Mur1 (198). All Afp encoding plasmids contain the operon-polarity suppressor (ops) sequence (5‘ GGCGGTAGCAT 3’), an element associated with the JUMPstart RNA sequence that improves downstream transcription. The AfpA region found on p1137 lacks this region, except for the 11bp ops sequence. Absence of this regulatory region on p1137 could lead to impaired transcription of the *afpA* operon or prevent the bacterial cell from lysing in order to release the AfpA particle. This could be tested by cloning the regulatory region into the p1337, upstream of the AfpA operon. Full alignment of the region is found in Supplementary Materials S.20

Table 7-7: Approximation of function of regulatory ORFs of Afp and Afpx based on BlastX hits.

Protein	Hit						Accession	
AnfA1	MULTISPECIES: transcriptional antiterminator NusG [*Enterobacteriaceae*]						WP_087638535.1	
148 AA residues	Length:	149	Range:	3 to 146	Eval:	2E-13		
	Ident:	48/150(32%)	Pos:	75/150(50%)	Gap:	11/150(7%)	Coverage:	97.0%
AnfA2	phage holin, lambda family [*Serratia entomophila*]						WP_010895797.1	
95 AA residues	Length:	95	Range:	1 to 95	Eval:	3E-60		
	Ident:	94/95(99%)	Pos:	94/95(98%)	Gap:	0/95(0%)	Coverage:	99.0%
Hypothetical	No functionally described hits							
95 AA residues								
enp1	lysis protein [*Serratia entomophila*]						WP_010895800.1	
153 AA residues	Length:	153	Range:	5 to 153	Eval:	1E-103		
	Ident:	145/149(97%)	Pos:	148/149(99%)	Gap:	0/149(0%)	Coverage:	97.3%
hol1	holin [*Citrobacter amalonaticus*]						WP_103780163.1	
67 AA residues	Length:	104	Range:	36 to 102	Eval:	5E-21		
	Ident:	42/67(63%)	Pos:	50/67(74%)	Gap:	0/67(0%)	Coverage:	62.7%
mur1	lysozyme [*Serratia entomophila*]						WP_010895802.1	
141 AA residues	Length:	141	Range:	1 to 141	Eval:	4E-98		
	Ident:	139/141(99%)	Pos:	140/141(99%)	Gap:	0/141(0%)	Coverage:	98.6%

Chapter 8 Novel acquired regions

The *S. entomophila* reference plasmid pADAP (83) was found to have two virulence determinants Sep and Afp as described in detail in Chapter 5 and Chapter 7. With the work undertaken on the *S. proteamaculans* strain AGR96X (69) as well as *S. proteamaculans* pU143 plasmid (83), it became evident that there were pADAP variants that were lacking either the Sep or Afp encoding regions respectively. With the work undertaken on the Sep region (90), there was also evidence of non-pADAP type plasmids encoding Sep. It was therefore assumed these islands can actively undergo transposition. As shown in Chapter 4 there is a high rate of sequence conservation across the pADAP-type backbone, as even the *S. marcescens* WVU-005-1 plasmid still maintained a large portion of DNA, with high similarity to the pADAP-type backbone (Section 4.11). This WVU-005-1 plasmid was found in a non-New Zealand isolate, which was obtained from a human clinical sample instead of insects, making it very unlikely both WVU-005-1 and pADAP share a recent common ancestor. If the backbone is considered a capturing device, it would be logical to assume that it could also capture novel regions other than Sep or Afp. Several of these acquisitions are described in this Chapter.

8.1 Novel region of unknown fuction

One of the most conserved novel regions identified in this study, is a GI of unknown function, designated as Region of Unknown Function (RUF). This region is approximately 39 Kb in size and is only found in *S. proteamaculans* based pADAP-type plasmids assessed in this study (Figure 8-1, Table 8-1, Supplementary Materials S.28).

As the DNA similarity matrix shows (Figure 8-2), this RUF region is highly conserved with an average of 99% sequence conservation. There are two distinct operons with a potential pathogenic property, the first is a cluster of fimbrial formation proteins followed by adhesin and exotoxin elements. The second region is a cluster of tightly packed hypothetical ORFs, that, until the reannotations of the RefSeq database, had hits to Type IV secretion system proteins. For the ease, these regions were designated Novel1 and Novel2 respectively.

The Novel2 region is also found independently in the two *S. proteamaculans* pADAP plasmids p1457 and p299, denoted as type (B) in Figure 8-1. Another variation of the region, denoted as type (C) in Figure 8-1, can be found in several plasmids, including three non-pADAP plasmids (Table 8-1). This variant region (C), has most of the Novel1 region, but has the Novel2 region inverted. No elements are found that were directly tied to the reorientation of Novel2 region in the (C) type.

All plasmids encoding RUF GI also encode for the Sep/Spp, but do not encode for an Afp PAI. However, not all Sep/Spp encoding plasmids contain this region: it is therefore not mutually exclusive to Sep/Spp encoding plasmids.

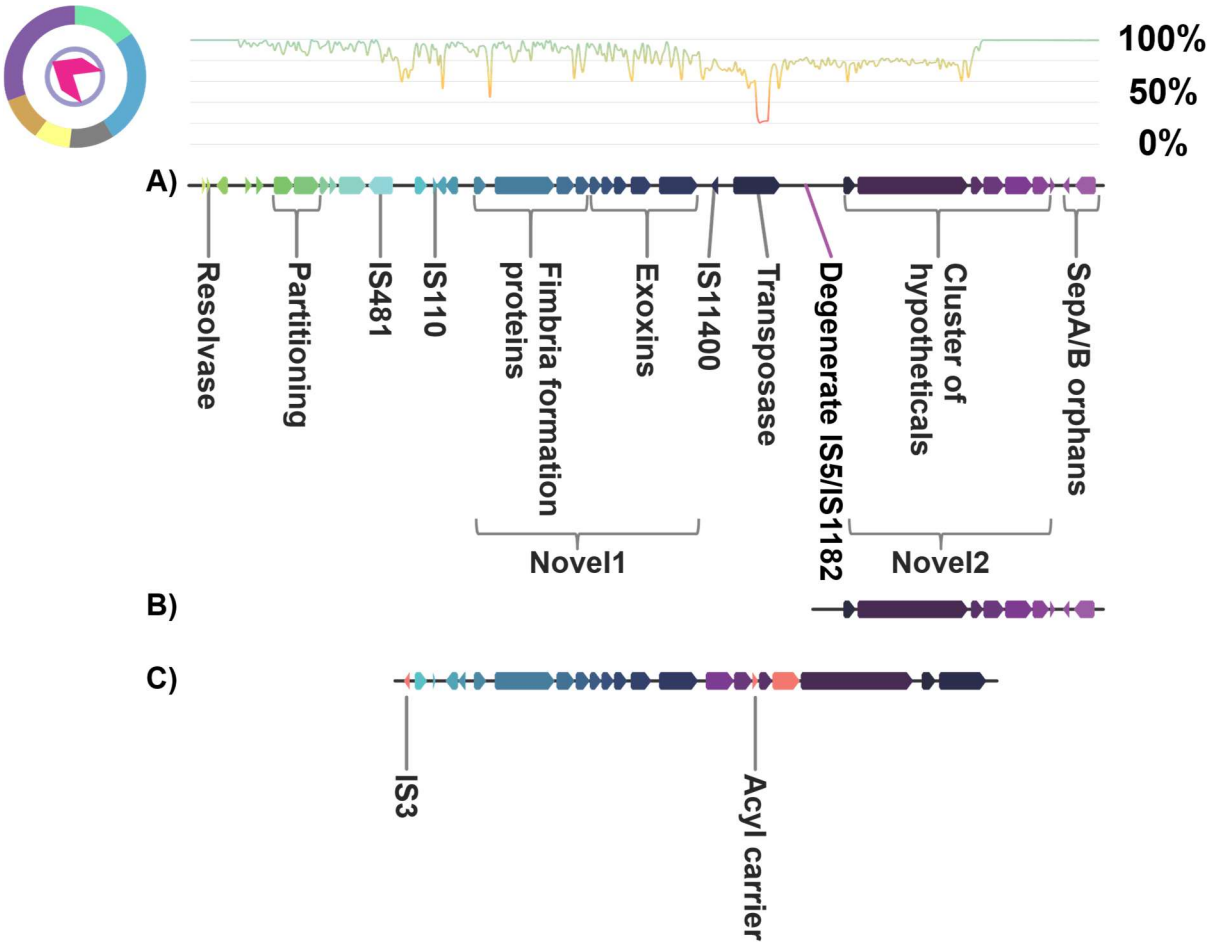


Figure 8-1: A schematic of the RUF GI. Top graph represents DNA similarity between all plasmids encoding for this region. There are several hypothetical ORFs, IS elements and transposases present. The (B) type, containing only Novel2, is found exclusively in p1457 and p299. The (C) type is mostly found in non-pADAP plasmids and features the fimbria/toxin region and a cluster of hypothetical ORFs (refer to Table 8-1 for isolates encoding A, B or C clusters).

Table 8-1: Distribution of isolates carrying each type of RUF GI.

Genotype	Species		Isolate
pADAP-type			
A	*Serratia*	*proteamaculans*	149, 336, 465, 1769, 1770 1772, 12a, 12d, 28F, D, M
B	*Serratia*	*proteamaculans*	299, 1457
Non-pADAP			
C	*Serratia*	*proteamaculans*	591, Ma3, Man4, S-prot-1

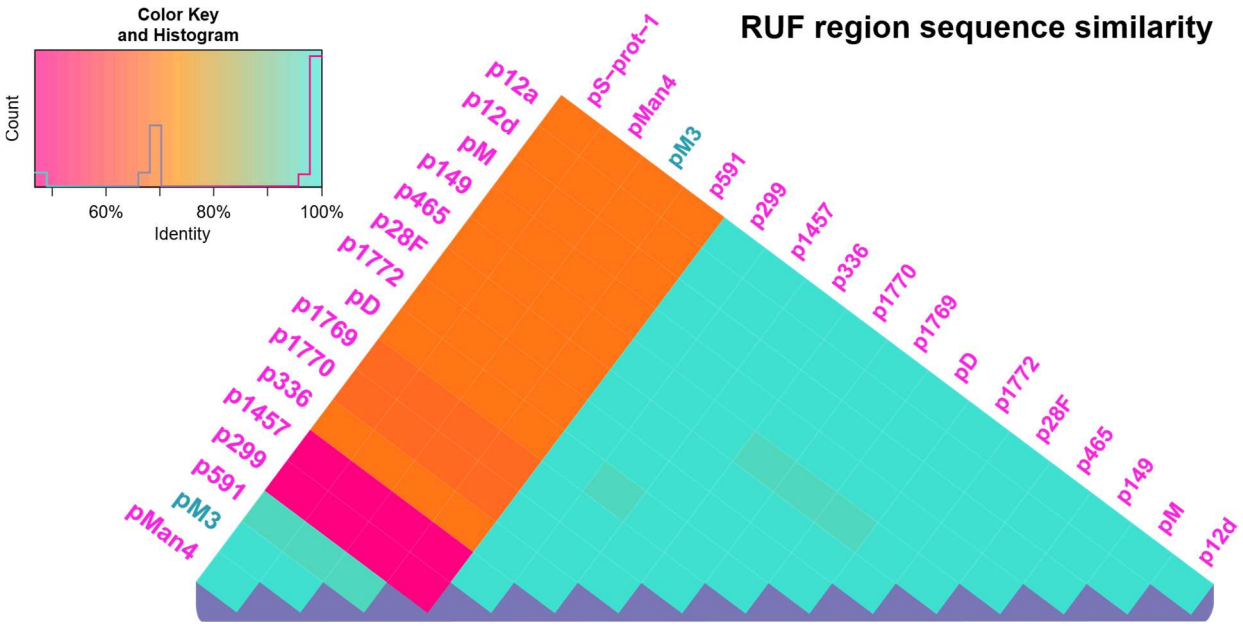


Figure 8-2: A DNA similarity matrix of the novel regions obtained from 17 different plasmids. Blue labelled strain names are *S. entomophila* and pink *S. proteamaculans*.

No BlastN hits were found for the RUF region, nor for the Novel1 and Novel2 sub-regions, meaning that no known orthologs of the island are currently present in the RefSeq database. As the two sub-regions Novel1 and Novel2 were found to be conserved in other pADAP and non-pADAP-type plasmids, they were deleted using targeted mutagenesis as outlined in Section 2.2.11 through 2.2.13, in order to determine whether these regions are associated with bioactivity against grass grub. Isolate 149 was the first isolate found in this study to contain the RUF on its pADAP-type plasmid, and therefore was used as the reference for this region. The regions were deleted using the method described in Section 2.2.11, 2.2.12 and 2.2.13. Besides independent targeted deletions of the two novel sub-regions of p149 RUF GI, the Sep region was also mutated using an inhouse vector 143SpRVpJP5603, deleting portions of the SepA and B region. All three p149 plasmid mutants were bioassayed as described in Section 2.4.1, together with the 149 WT and A1MO2 isolate, as well as the appropriate positive and negative controls. The results are shown in Table 8-2.

Table 8-2: Bioassay results of *S. proteamaculans* p149 mutants compared to the *S. proteamaculans* 149 and *S. entomophila* A1MO2 WT strains.

Strain	Disease%	Mortality%	Affected%
Control	2.4 ± 1.7	1.2 ± 1.2	3.6 ± 2.0
A1M02	83.3 ± 4.1	3.6 ± 2.0	86.9 ± 3.7
149	13.9 ± 5.8	11.1 ± 5.3	25.0 ± 7.3
149ΔSprV	0 ± 0.0	0 ± 0	0 ± 0
149Δnovel1	8.3 ± 3.0	0 ± 0	8.3 ± 3
149Δnovel2	8.3 ± 3.0	0 ± 0	8.3 ± 3

The bioassay results show that the 149 WT itself is not highly bioactive in grass grub, with only a basal level of activity observed at day 12. The mutant of the Spp region shows complete loss of any residual bioactivity. The knockout of the Novel1 and Novel2 regions resulted in minor reduction of bioactivity compared to wildtype, and therefore it is likely that these regions themselves do not contribute to the insecticidal activity in grass grub, and that Spp is the main virulence determinant on the p149 plasmid. The 149 strain was also tested on three other beetle species found in NZ to find potential alternate hosts that the novel regions could have effects on and these results are shown in Table 8-3.

Table 8-3: Bioassay results of isolate *S. proteamaculans* 149 against several non-Grass grub beetle species.

Insect	Disease%	Mortality%	Affected%
Red Headed Chafer	0 ± 0.0	0 ± 0.0	0.0 ± 0.0
Manuka beetle	0 ± 0.0	10 ± 10.0	10 ± 10
Tasmanian Grass Grub	0 ± 0.0	0 ± 0.0	0.0 ± 0.0

The bioassays against larvae of several other New Zealand based beetle species showed no significant bioactivity (Table 8-3). It is therefore likely that either the novel region is associated with a yet to be determined organism or has yet to be determined alternate. Full functional description of the ORFs can be found in Supplementary materials S.28. Full alignment of the region is found in Supplementary Materials S.20

8.2 Region downstream of Sep

As mentioned in the previous section, most Sep encoding pADAP-type plasmids are flanked by RUF, however not all Sep encoding plasmid transcribe for this GI. Those plasmids that encode for the Sep PAI, but not RUF, do not have a consistent shared downstream region, instead are often flanked by one or two transposable elements at the 3' side before transitioning back into the pADAP backbone region of replication described in Section 4.4.

8.3 Afp downstream regions of divergence

As outlined in the Chapter 7, there are three types of Afp PAIs found in the plasmids assessed in this study. These three are the Afp, AfpX and AfpA PAIs. The Afp and AfpX variants have a regulatory region upstream as described in Section 7.4, whereas the AfpA containing isolate 1137 has no upstream associated region, except for the demarcation island discussed in Section 4.12. Downstream these three Afp variants also have very divergent regions as described in the following sub-sections.

8.3.1 pADAP Afp downstream region

The pADAP plasmids found in the *S. entomophila*, all share a nucleotide similarity throughout the entire plasmid including the backbone, Sep and Afp regions. The p142 and p145, were the only two *S. proteamaculans* plasmids found to bear the *S. entomophila* pADAP. Each of these pADAP plasmids contain an int1 protein and several Sea ORFs downstream of the Afp PAI (Figure 8-3, Table 8-4).

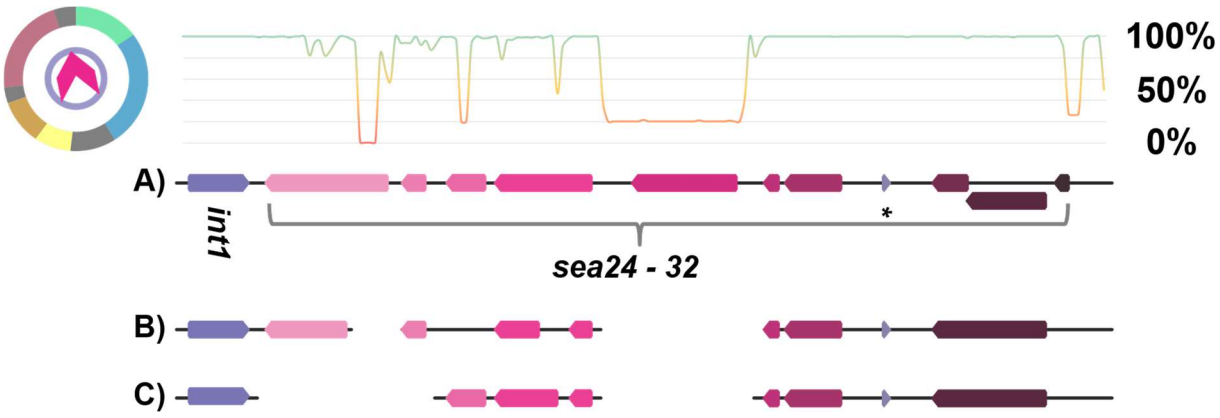


Figure 8-3: The downstream Afp region of pADAP plasmids approximately 9 Kb in size. Refer to Table 8-4 for isolates encoding type A, B or C. Top graph represents DNA similarity between all plasmids encoding for this region.

Table 8-4: Distribution of isolates carrying each type of downstream Afp region.

Type[1]	Species		Isolate
A	*Serratia*	*entomophila*	345, 398, Diarr, pADAP
B		*entomophila*	210, 1100, Moraki_2
		proteamaculans	145
C		*entomophila*	158, 176, 626
		proteamaculans	142

[1] Type defined in Figure 8-3

Table 8-5: BlastX hits of sea24-32. These elements seem to mostly be transposable elements with little effect of Afp production or release.

Original annotation	New annotation	Query Cover	E value	Identity	Accession
sea24	ISNCY family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895822.1
sea25	ISNCY family transposase [*Serratia* sp. S4]	81%	1.00E-21	91.04%	WP_083906151.1
sea26	ISNCY family transposase [*Serratia entomophila*]	99%	8.00E-63	100.00%	WP_010895822.1
sea27	ISNCY family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895825.1
sea28	ISNCY family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895826.1
Sea29	nuclear transport factor 2 family protein [*Serratia* sp. BW106]	57%	7.00E-11	84.85%	WP_099062043.1
sea30	DUF2913 family protein [*Serratia entomophila*]	99%	1.00E-140	99.48%	WP_010895828.1
sea31	latent-transforming growth factor beta-binding protein 1-like [*Meriones unguiculatus*]	67%	0.84	32.18%	XP_021508259.1
sea32	5’/3’-nucleotidase SurE [*Halovivax asiaticus*]	34%	5.2	26.67%	WP_007698437.1

These *sea* ORFs found in the downstream Afp regions of pADAP plasmids mostly encode for IS Not Classified Yet (ISNCY) family transposases (Table 8-5). Further analysis of this cluster of ISNCY transposases is provided in Section 9.2.3.

8.3.2 Colicin associated region

Five plasmids assessed in this study were found to contain a highly conserved ~6 Kb region that encodes a Colicin based antimicrobial / post-segregational killing system (PSKS) (Figure 8-4, Section 11.2).

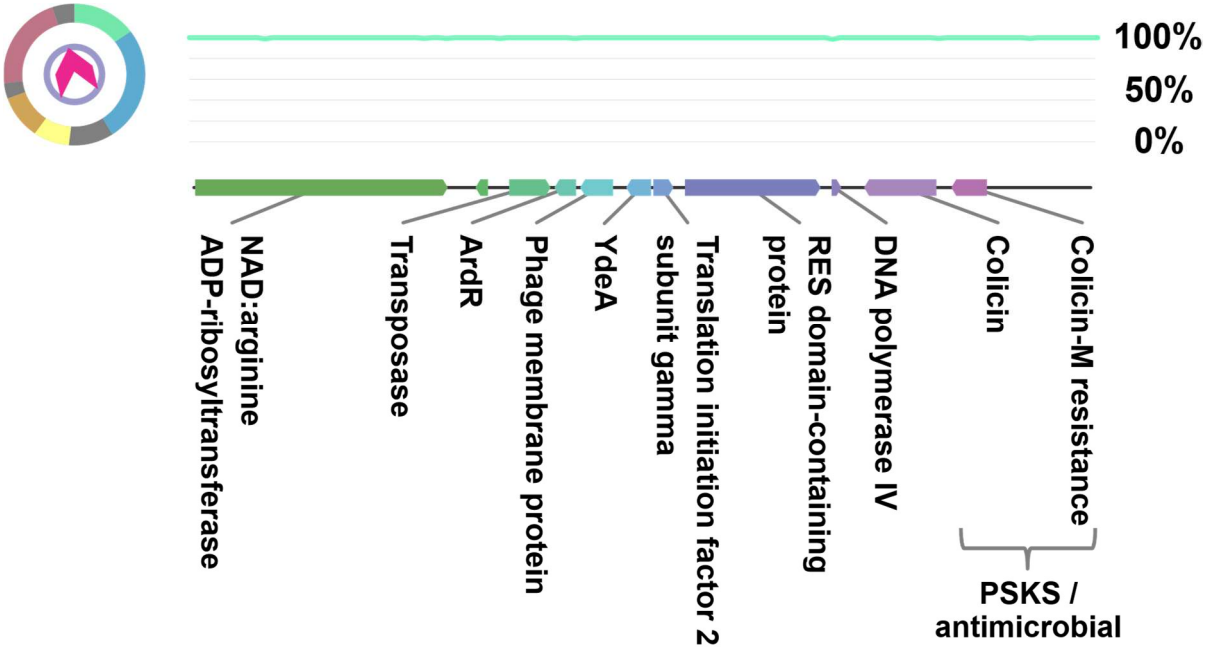


Figure 8-4: The ~6 Kb colicin associated region, located 3' of the Afp encoding region of pADAP-type plasmids found in isolates 4, 163, 1129, 1A and SpF. Top graph represents DNA similarity between all plasmids encoding for this region.

This region contains several genes of interest, not documented previously on pADAP-type plasmids. These include an anti-restriction and anti-modification protein ArdR (299) ortholog, a protein that prevents genetic modification by foreign DNA. Another ORF unique to the pADAP system is a YdeA (337) ortholog. The YdeA is part of the major facilitator superfamily (MFS), a family of proteins that transport molecules across the OM through use of protons. YdeA is directly correlated to L-Arabinose and Isopropyl-β-d-Thiogalactopyranoside export.

Table 8-6: Closest functional BlastX hits for ORFs present in the 1129-like downstream Afp region.

Original annotation	New annotation	Query Cover	E value	Identity	Accession
Afp17X	NAD:arginine ADP-ribosyltransferase [*Photorhabdus australis*]	96.0%	0	67.7%	OCQ51405.1
NA	Uncharacterised protein [*Yersinia enterocolitica*]	86.0%	5E-19	90.7%	CFQ34327.1
NA	transposase for insertion element IS1222 [*Yersinia enterocolitica* subsp. Enterocolitica WA-314]	66.0%	2E-63	95.1%	EKA28337.1
NA	antirestriction protein ArdR [*Salmonella enterica*]	98.0%	3E-25	62.8%	EAX5070673.1
NA	putative phage membrane protein [*Salmonella enterica* subsp. Enterica]	91.0%	2E-43	61.9%	VEA23963.1
NA	YdeA protein [*Shigella sonnei*]	72.0%	7E-19	58.0%	ARB02219.1
NA	MULTISPECIES: hypothetical protein [Enterobacteriaceae]	87.0%	2E-19	63.5%	WP_017901039.1
NA	RES family NAD+ phosphorylase [*Serratia* sp. S4]	92.0%	0	90.2%	WP_017891074.1
NA	lipid II-degrading bacteriocin [*Pantoea ananatis*]	97.0%	1E-72	44.9%	WP_110286972.1
NA	colicin M resistance protein [*Pantoea* sp. Sc1]	87.0%	3E-14	32.5%	EIB96703.1

Table 8-7: Closest functional BlastX hits for ORFs present in the downstream AfpX region.

New annotation	Query Cover	E value	Identity	Accession
IS3-like element ISSen4 family transposase [*Citrobacter freundii*]	72.0%	9E-85	80.0%	WP_000928387.1
MULTISPECIES: hypothetical protein [*Klebsiella*]	99.0%	4E-163	54.5%	WP_101880745.1
MULTISPECIES: hypothetical protein [*Enterobacteriaceae*]	82.0%	2E-12	50.9%	WP_020804422.1
MULTISPECIES: hypothetical protein [*Serratia*]	84.0%	1E-21	62.3%	WP_060452169.1
Restriction endonuclease [*Pectobacterium carotovorum*]	99.0%	0	93.8%	ALG88642.1
MULTISPECIES: type II toxin-antitoxin system VapC family toxin [*Serratia*]	99.0%	5E-78	94.9%	WP_048796372.1
AbrB/MazE/SpoVT family DNA-binding domain-containing protein [*Kosakonia* sp. WCHEs120001]	98.0%	6E-44	93.4%	WP_131413711.1
histidinol-phosphatase [*Salmonella enterica* subsp. Enterica]	99.0%	0	93.6%	EBG3215765.1
DNA-binding protein [*Citrobacter* sp. Wls710]	97.0%	0	59.6%	WP_137349397.1
hypothetical protein [*Serratia plymuthica*]	52.0%	6E-25	69.0%	WP_037429921.1

Two ORFs encoded on this region are a colicin ortholog and a colicin M resistance protein ortholog (as labelled in Figure 8-4). These two proteins form a bactericidal TA system that can act as an antimicrobial towards non anti-toxin bearing cells, as well as a PSKS that helps stabilize the plasmid within a population. More on this particular TA is outlined in Section 11.2.

8.3.3 AfpX associated region

Four of the AfpX encoding pADAP-type plasmids, p20093, pLC, pAGR96X and pSprot5, contain a ~13 Kb downstream region that also transcribes a PSKS (Figure 8-5, Table 8-7), that differs to that of the colicins mentioned in the previous section. The VapBC-like TA is a bacteriostatic TA, which through the cleavage of tRNAs, can force cells into a metabolic persister state (148). Further detail on VapBC TA is presented in Section 11.3. Of interest, a ~4.6 Kb homologous portion of the region was found on a plasmid from a French *S. entomophila* isolate 220 (Figure 8-5,). This p220 plasmid only shares the VapBC TA, histidinol-phosphatase and a truncated portion of the restriction endonuclease found on the downstream AfpX region of the *S. proteamaculans* plasmid, pAGR96X.

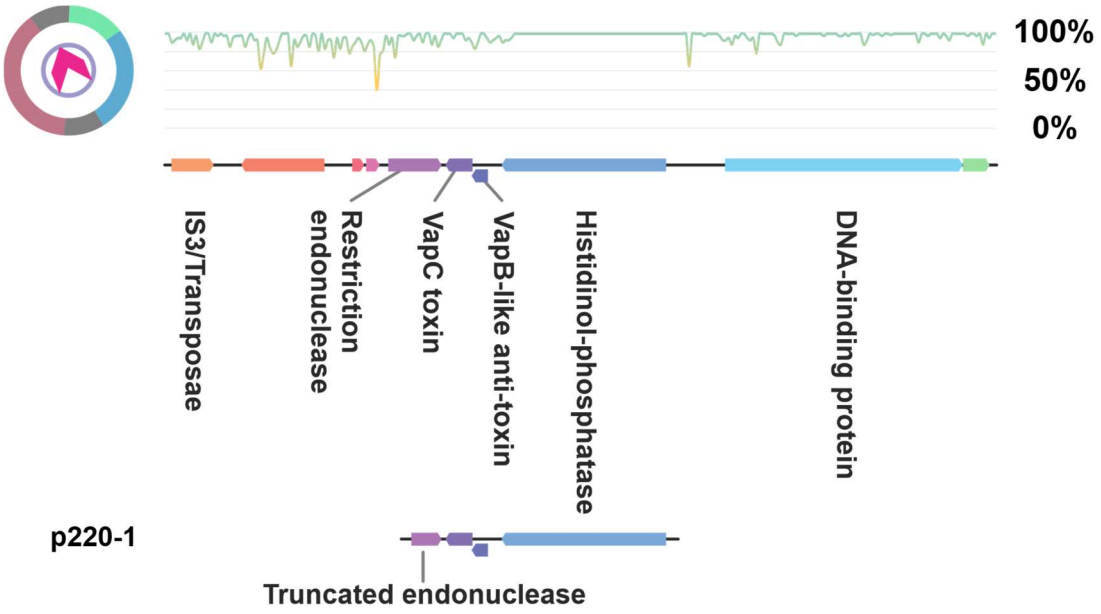


Figure 8-5: The ~13 Kb AfpX associated region of, p20093, pLC, pAGR96X and pSprot5 located 3' of AfpX encodes for a VapBC TA. Similar region identified in the p220-1 plasmid. Top graph represents DNA similarity between all plasmids encoding for this region.

8.3.4 AfpA associated region

The *S. proteamaculans* pADAP-type plasmid from isolate 1137 contains a distinct ~13 Kb region, located 3' of the AfpA PAI (Figure 8-6), previously outlined in Section 7.3. This AfpA associated region consists of several ORFs also present on the downstream AfpX region, as well as several distinct ORFs encoding a protease, transposase and SOS-response associated protein.

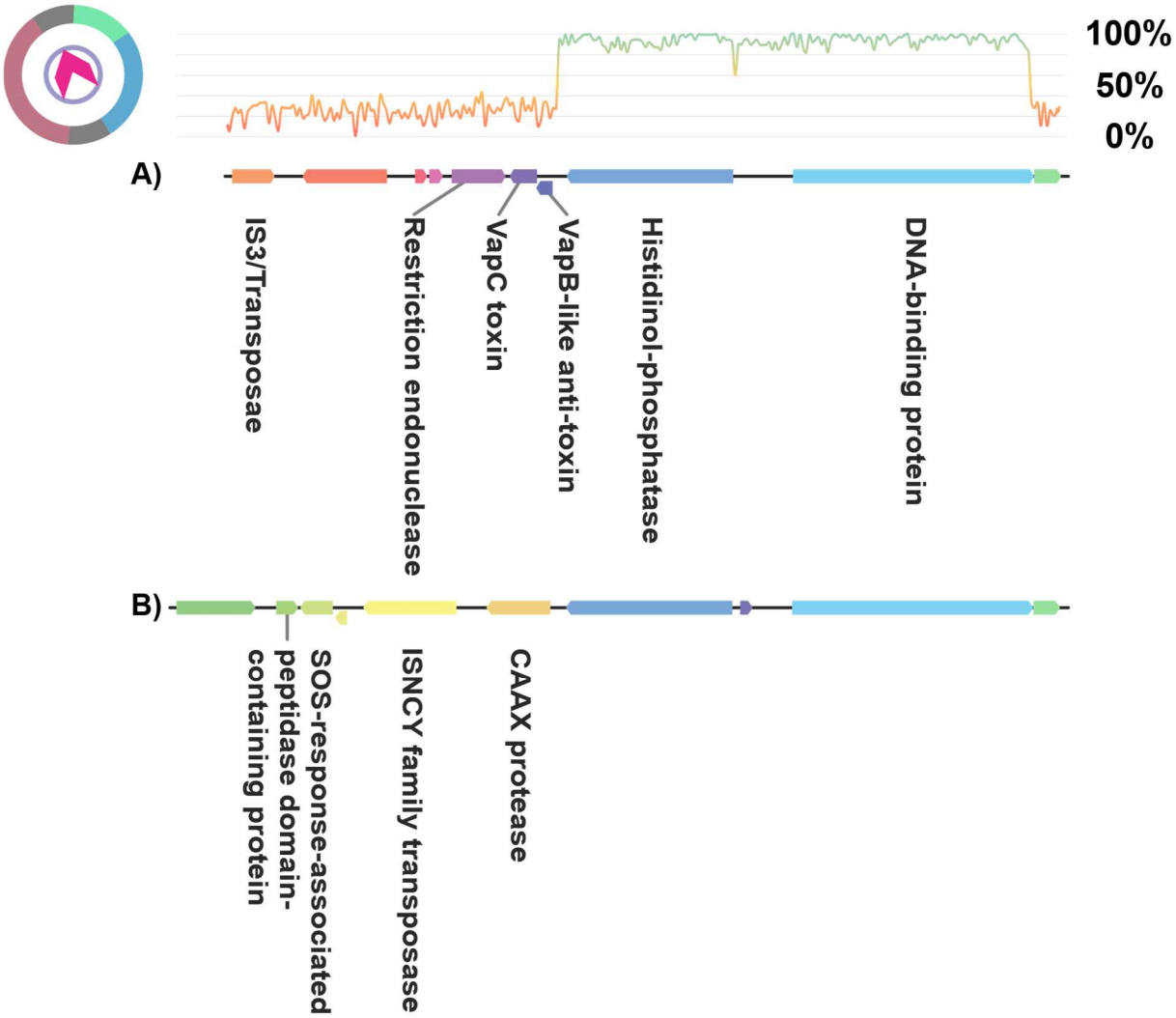
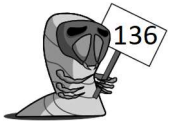


Figure 8-6: The ~13 Kb AfpA associated region, located 3' of the p1137 encoding GI. The p1137 AfpA associated region (B) compared to the AfpX associated region of pAGR96X (A). Top graph represents DNA similarity between pAGR96X and p1137 plasmids encoding for this region.

Table 8-8: Closest functional BlastX hits for ORFs present in the downstream AfpA region.

New annotation	Size (AA)	Query Cover	E value	Identity	Accession
MULTISPECIES: hypothetical protein [*Pantoea*]	390	99.0%	1E-98	42.9%	WP_105099763.1
SOS response-associated peptidase [*Salmonella enterica*]	116	31.0%	1E-06	64.9%	EAZ4751223.1
MULTISPECIES: hypothetical protein [Enterobacteriaceae]	160	99.0%	1E-93	98.1%	WP_022631300.1
MULTISPECIES: hypothetical protein [Enterobacteriaceae]	69	98.0%	3E-38	97.1%	WP_031280849.1
ISNCY family transposase [*Klebsiella quasipneumoniae* subsp. Quasipneumoniae]	456	99.0%	0	92.5%	OVY28499.1
CAAX protease [*Serratia marcescens*]	333	99.0%	0	93.7%	WP_149570805.1
MULTISPECIES: histidinol-phosphatase [Enterobacteriaceae]	874	99.0%	0	92.6%	WP_063097007.1
hypothetical protein ACN56_05890 [*Escherichia coli*]	72	79.0%	0.009	42.1%	OKS95488.1
DNA-binding protein [*Salmonella enterica* subsp. Diarizonae]	1264	97.0%	0	59.5%	EBF4785257.1



8.4 **pPuna18 plant associated island**

One of the outlier plasmids analysed in this study is a pADAP-type plasmid found in the *S. proteamaculans* Puna18, named for the location where it was found, Punakaiki, on the West Coast of New Zealand's South Island. This plasmid encodes the pADAP-type backbone, previously outlined in Section 4.3, and shares 94.1% nucleotide similarity to the *S. entomophila* pADAP backbones as previously discussed in Section 4.6, and only 85.9% to the *S. proteamaculans* pAGR96X. The pPuna18 plasmid does not contain the ~3.9Kb insertion in between the *traG* and *trbC* ORFs located on the backbone, previously described in Section 4.9, that is associated with the majority of all *S. proteamaculans* pADAP-type plasmids analysed in this study. The absence of this insertion makes the pPuna18 plasmid backbone more similar to those of pADAP and WVU-005-01 and might provide additional evidence that this insertion has been acquired at a later time point. Therefore it is likely that the pADAP and pPuna18 backbones are more indicative of the ancestral *Serratia* associated capturing apparatus.

Importantly, the point of demarcation, previously discussed in Section 4.12, is where the pPuna18 starts to diverge from other pADAP-type plasmids. Downstream of the demarcation region, on the 3' side, the pPuna18 does not encode for the pADAP associated PAIs and GIs such as Sef (Chapter 6), Sep (Chapter 5) or Afp (Chapter 7), or the newly identified RUF GI discussed in Section 8.1, or other elements that may relate to an insect association. The novel pPuna18 region, presented in Figure 8-7, encodes for numerous metabolite related proteins, most notably a large cluster of 12 Nif ORFs. Nif genes are involved with fixation of nitrogen from freely available nitrogen (N_2) to ammonia (NH_3) which is then convertible to ammonium (NH_4), which can be taken up by plants for further metabolization (338, 339). Other elements found on this novel region are two ORFs highly homologous to Flavodoxin proteins. Flavodoxins facilitate electron transfer to nitrogenase (340). There also appear to be three separate operons with ABC transporter encoding ORFs (341). ABC stands for ATP binding cassette, and members of the ABC transporter family of proteins which are implicated in the transport of a range of substrates (342, 343). Full annotation of the novel pPuna18 ORFs is shown in Figure 8-7 and in Supplementary Materials S.30.

Several other metabolite-associated ORFs are encoded on the pPuna18, but the majority appear to be involved in plant associated metabolites. It's possible that the *S. proteamaculans* isolate Puna18 has used the *Serratia* associated capturing apparatus to co-evolve with plant species, as the *S. entomophila* pADAP bearing A1MO2 type-strain appears to have co-evolved with larvae of the *C. giveni*. Based on shared nucleotide similarity and general orientation of the backbone of pPuna18 compared to other pADAP-type plasmids, it is likely all these plasmids started with one shared capturing apparatus and diverged based on target organisms each isolate cohabitates with.

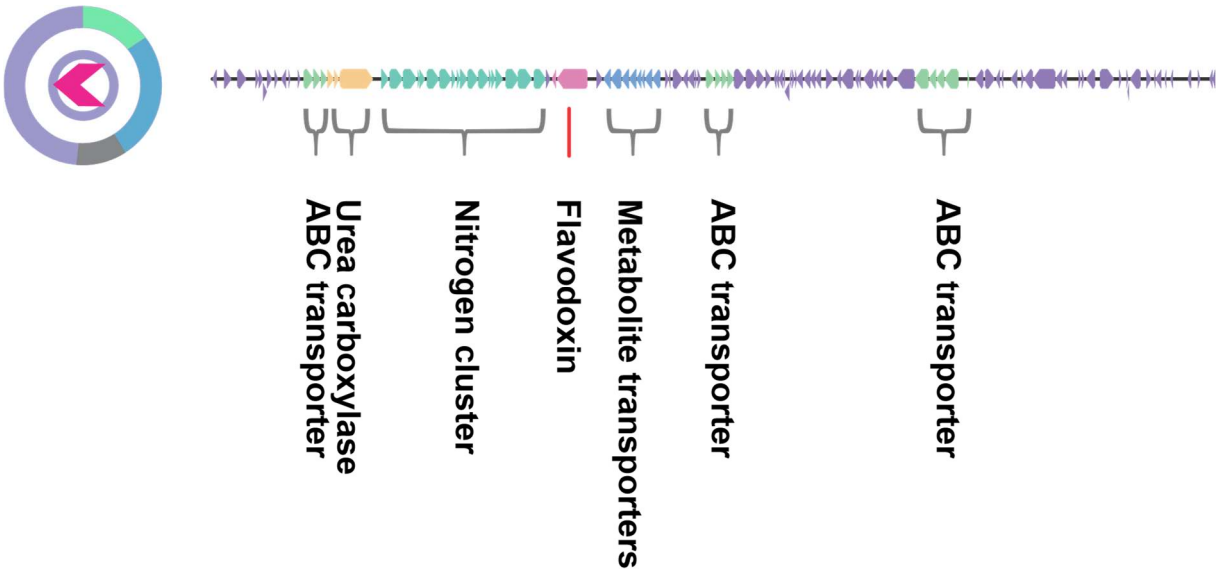


Figure 8-7: The ~114 Kb novel region identified in pPuna18. The region encodes for 116 proteins, including a nitrogen fixing cluster, a range of transposable elements and metabolite associated proteins. Detailed annotations the pPuna18 plant associated island can be found in Supplementary Materials S.30.

The *S. proteamaculans* isolate Puna18, carrying the pPuna18 STAMP which encodes for a ~24 Kb nitrogen fixation cluster, orthologous genes of which, encoded by the *Sinorhizobium meliloti* mega-plasmids pSymA and pSymB, have facilitate microbe-plant interactions in rhizobia (344). The pPuna18 *nif* cluster showed 90.4% nucleotide sequence identify to the plasmid pRahaq202 (GenBank accession no. CP003246.1) found in *Rahnella aquatilis*, a plant symbiont associated with growth promotion (345). These findings suggest that *S. proteamaculans* isolate Puna18 most likely interact with plants rather than *C. giveni* larvae.

8.5 Spb Bacteriophage

Two plasmids assessed in this study were found to carry a putative bacteriophage. The bacteriophage associated region (Figure 8-8), termed *Serratia proteamaculans* bacteriophage (Spb) is approximately ~ 37 Kb in size and is found in the pADAP-type plasmid p465 and the non-pADAP-type plasmid p591. The Spb regions of p465 and p591 share an 88.2% nucleotide sequence similarity.

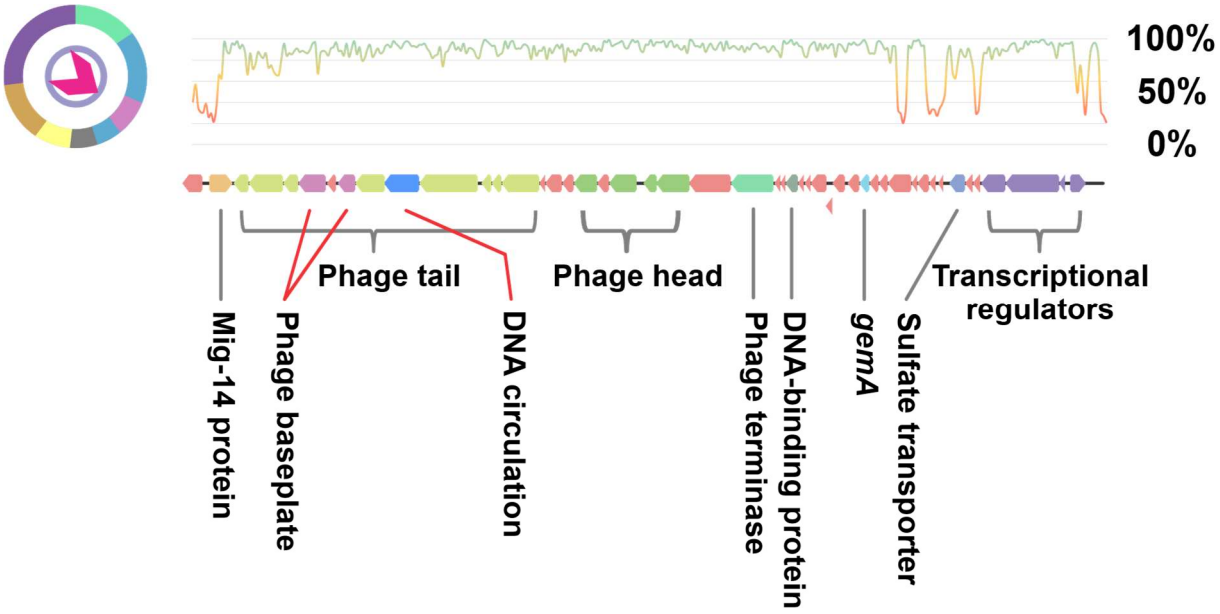


Figure 8-8: The ~37 Kb Spb bacteriophage associated region found in *S. proteamaculans* pADAP-type plasmid p465 and the non-pADAP-type plasmid p591. The region encodes for 48 proteins, of which the majority have phage associated orthologs in the RefSeq database. Top graph represents DNA similarity between all plasmids encoding for this region. Refer to Supplementary Materials S.31 for detailed annotations.

The putative phage particle has been designated Spb, for *Serratia proteamaculans* Bacteriophage. The region contains eight phage tail encoding ORFs, two baseplate-encoding ORFs as well as a third ORF belonging to a baseplate protein family (*spb26*, not depicted in the figure), four phage head associated ORFs, a DNA circulation protein and a Phage terminase encoding ORF. As all units of a bacteriophage appear to be encoded on the Spb island it is likely this is an active, functional bacteriophage. Further evidence of this is its presence on two distinct plasmids. The region itself does not encode of a lysis cassette.

There are several phages that replicate in a bacterial host through the formation of a small circular replicon, i.e. plasmid, such as the such as the Bacteriophage P1 found in *E. coli* (346) and the Bacteriophage LE1 found in *Leptospira biflexa* (347). Through a literature search this is the first example of a putative functional plasmid encoded bacteriophage. There are man-made replicons consisting of a plasmid ori and a phage genome, such as the pEMBL plasmid family (348).

These types of replicons are referred to as phagemids (349). Even though bacteriophage insertion and replication from a plasmid source is therefore plausible, no literature was uncovered discussing plasmid-based bacteriophages.

The Spb region contains several hypothetical ORFs, DUF encoding ORFs and a cluster of transcriptional regulators. It is possible that this additional region is accidental carryover from the DNA packaging, or that it's a later acquisition with undefined phage regulatory purposes. This potential of it being a later acquisition is further supported by a sudden drop in GC% compared to that of the majority of the bacteriophage encoding ORFs (Figure 8-9). As shown, the GC% does vary across the region, with most of the phage encoding ORFs having an average GC% of >60%.

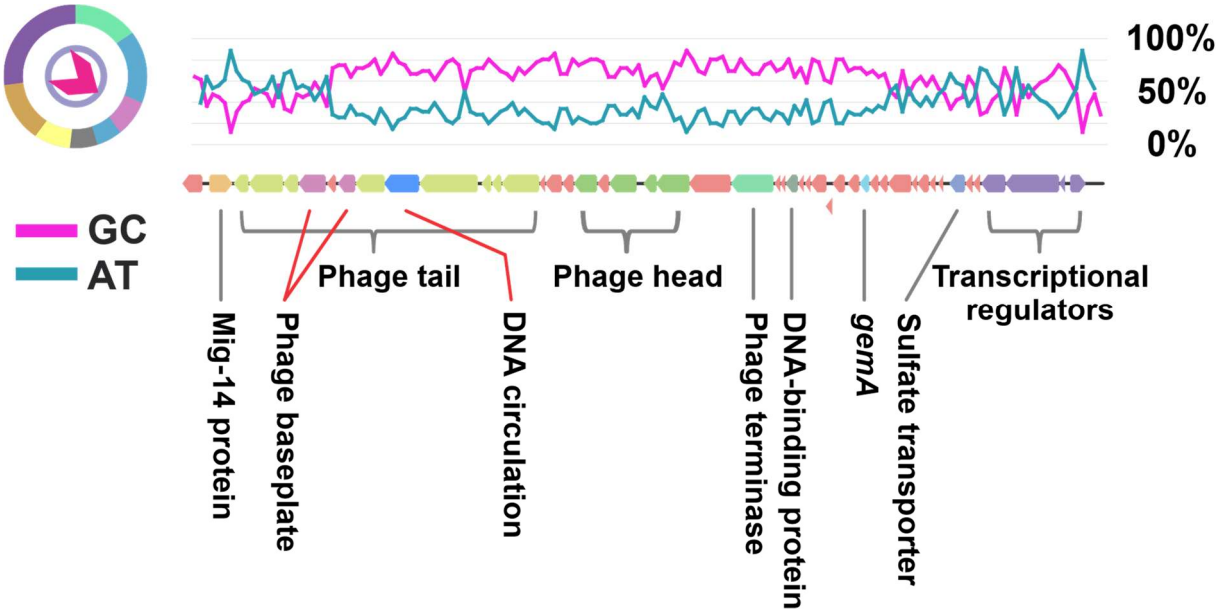


Figure 8-9: The GC% of the ~37 Kb Spb bacteriophage associated region found in *S. proteamaculans* pADAP-type plasmid p465

Full list of BlastX hits for each ORF in the region are present in Supplementary Materials Appendix AS.31.

The GC content of the Spb region shows significantly higher GC% than the surrounding area (Figure 8-10), approximately 55.4% for the *S. proteamaculans* p465 plasmid and 55.5% for the *S. proteamaculans* p591 (not shown in Figure 8-10), compared to the GC% of the remainder of these plasmids which is 49.0% and 49,9% respectively.

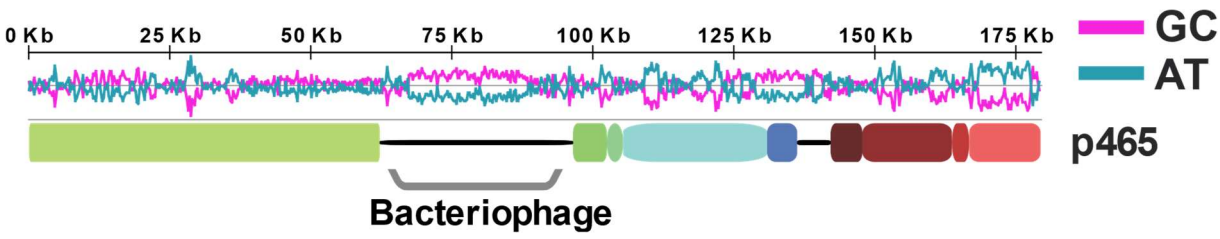


Figure 8-10: GC content of bacteriophage region found on p465 compared to entire plasmid.

No scar or insertion site has been observed for the region as it inserted intergenic in the PilO ORF for p465. AT-rich spikes are present at the edges of the inserted bacteriophage, further suggesting active transposition events. Using RepFind, a tool to identify repeated motifs in a submitted DNA sequence (350), no significant repeated elements were found flanking the conserved bacteriophage encoding region, nor any shared transposable elements other than the identified elements presented in Figure 8-8.

To define if Spb encodes a functional bacteriophage, two genetic regions of the Spb, I) a portion of head associated ORFs and II) a portion of the tail associated ORFS, were independently as well as consecutively deleted through cloning and targeted mutagenesis as described in Section 2.2.11 through 2.2.13. These would serve as control for TEM post MitC induction. In addition to this, to omit the feasibility of the chromosomally encoded phages being expressed, the chromosome sequence was assessed, where no phage encoding regions were identified. Several attempts were made to purify the bacteriophage as described in Section 2.2.18, and record electron micrographs of it using TEM as described in Section 2.2.19, but no phage like entities were observed.

8.6 WVU-005-1 novel region

The WVU-005-1 plasmid derived from a clinical *S. marcescens* sample, contains a pADAP-type plasmid that shares the majority of the pADAP backbone, as described in Section 4.11. The only point of divergence in the backbone is the origin of replication not being similar to that of pADAP, and instead the replication genes are present in the region described as the demarcation region as previously outlined in 4.11 and Section 4.12. The WVU-005-1 plasmid does not carry any of the known or newly identified pADAP virulence associated regions such as Sep or Afp, or any accessory regions identified through the course of this study as outlined in this chapter. Instead, the WVU-005-1 plasmid encodes a range of secondary metabolic enzymes.

As this US based isolate is unlikely to have had any recent interaction with the New Zealand based isolates assessed in this study, it is safe to assume the ancestral *Serratia* associated capturing apparatus (i.e. what has become the current pADAP backbone) was very similar.

It is likely a distant geological split as well as time spent occupying different niches has led to the loss of its repA and using a replication protein more in tune with *S. marcescens*.

Full annotation of the novel region found on WVU-005-1 is provided in Supplementary Materials S.32.

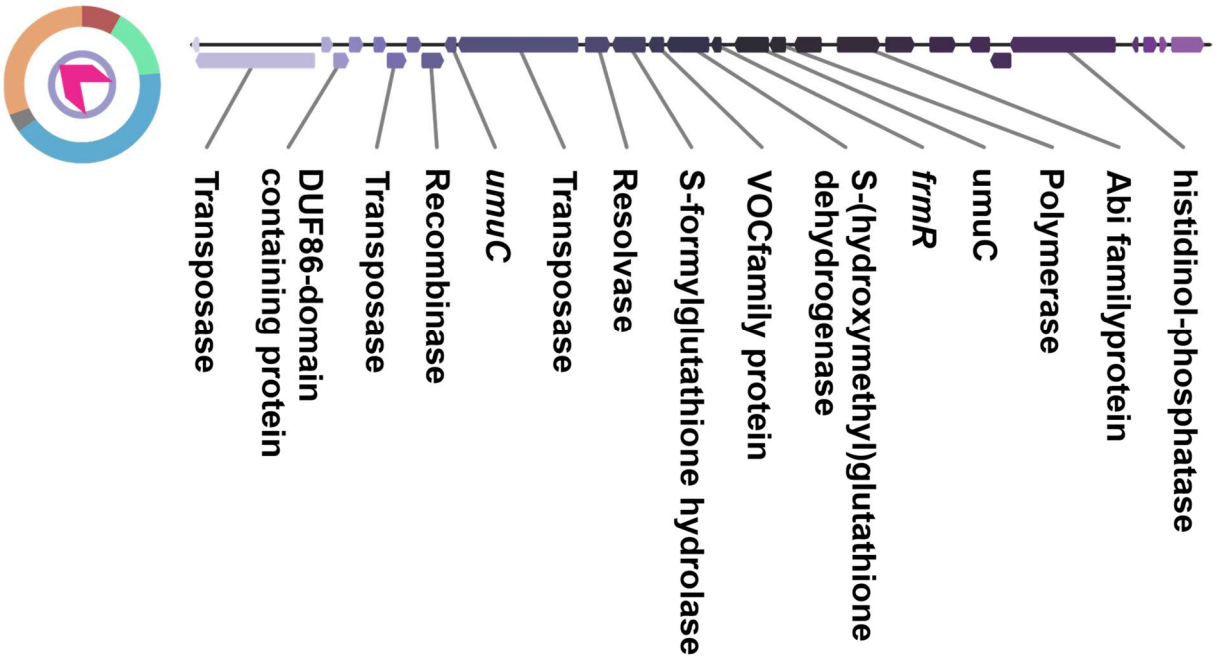


Figure 8-11: The ~25 Kb novel region identified in WVU-005-1. The region encodes for 29 proteins, of which the majority are transposable elements or metabolite associated proteins. Full annotations of this region can be found in Supplementary Materials S.32.

Chapter 9 MGE elements and HGT Drivers

In a paper by Hurst et al. published in 2003 (192), possible horizontally acquired regions and elements were identified in pADAP. In 2011 the complete pADAP plasmid was sequenced and all its features annotated (83). The putative PAI found on pADAP such as the Sep and Afp encoding regions included regions with high DNA identity to known IS elements. Other elements found were designated as integrases (int) and transposases. Apart from the annotated elements the plasmid also contained numerous predicted ORFs that were designated as *S. entomophila* annotation (*sea*) followed by a number, as at the time no known homologs were present in the NCBI database to infer their function. Through this project, several of these pADAP *sea* ORFs have now been shown to encode for some sort of MGE or HGT driver. A full list of the reannotation of the pADAP *sea* ORFs can be found in Supplementary Materials S.8. This chapter goes further into the detail of MGE found on pADAP.

9.1 Integrases or not?

The annotation of pADAP revealed numerous MGEs and hallmarks of horizontally acquired regions. In addition to this, pADAP itself is also a major driver of HGT and, due to the presence of a conjugative pore and sex pili, can be considered a mobile genetic element. Several pADAP-related HGT drivers of genetic regions have already been documented in previous studies undertaken by Hurst et al. (83), two of which were described as integrases, designated *int1* and *int2*. However, recent annotation has found that they are in fact not integrases. Int1 is a member of the IS3 family transposase. Transposases belonging to the IS3 family such as IS911, the most well-known member of the IS3 family member, are simple transposable elements. They transcribe their own transposase that helps with transposition of the region by circularizing it and cleaving it out of the region (351, 352). These MGEs are often implicated in the rapid genetic changes and drivers of genome streamlining through inactivation of unnecessary genes by intergenic insertion (353). Upon insertion into a new region, short imperfect terminal inverted repeat sequences (IRs) are often left by IS3 members. In the case of the *int1* ORF a degenerate 7 bp repeat with the consensus sequence (5' GCGGATG 3') was identified, flanking the *int1* ORF as seen in Figure 9-1. Four copies of this motif are present, twice outside of the int1 ORF and twice internal to the ORF itself.

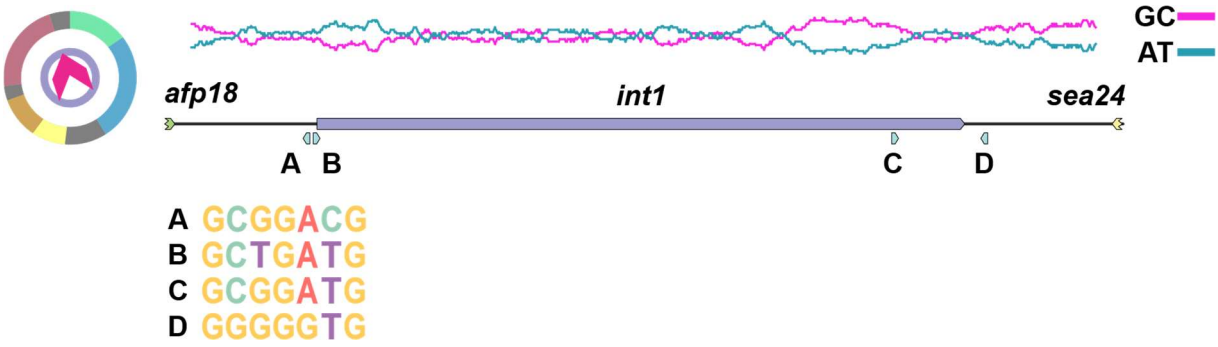


Figure 9-1: The region surrounding *int1*. Int1 is an IS3 family transposable element. The A-D features observed in pADAP, denote short imperfect terminal IRs often associated with these types of transposable elements (353).

The *int2* element appears to be an integral part of the backbone of pADAP type plasmids as it is highly conserved and present in all pADAP plasmids, even the outlier plasmids with non-insecticidal features such as pPuna18. The *int2* ORF shares homology with several resolvases in *S. marcescens*, *Edwardsiella tarda* and several other prokaryotes. The Int2 element contains a conserved C-terminal DNA breaking-re-joining enzyme domain, which is highly homologous to a putative recombinase found in the *orf18* of the *Yersinia enterocolitica* F plasmid (83). Upstream there are no conserved element other than the point of demarcation of the pADAP plasmid origin as described in Section 4.4. Downstream the *Int2* gene is flanked by an *atoX* and *staB* gene and followed by the origin of replication and the plasmid backbone.

9.2 IS elements

There were two remnant IS elements on the pADAP plasmids that were designated as IS *S. entomophila* 1 (Isse1) by Hurst et al. (83). The Isse1 element contains homologous areas to the SamB UV protection protein encoded by *Salmonella enterica* subsp. enterica serovar Typhi. Isse2, which is an IS66-like element, shows an >80% AA similarity with IS elements in *E. coli* and *Shigella* species (192). Both IS elements reside downstream of the *sepC* ORF. Several pADAP variants, including the pU143, previously discussed by Hurst et al. (83), are missing the Isse2, IS91-like elements and the *afp* gene cluster. Although this absence was suggested to be correlated with acquisition of the Afp region by pADAP through HGT, it is highly unlikely IS elements can facilitate transpositions of islands this size. Apart from the two annotated IS elements, several *sea* ORFs have been re-annotated as IS elements. Revised annotations are listed in Supplementary Materials S.8.

9.2.1 Sea14 and 15

Both *sea14* and *sea15*, residing in the demarcation region outlined in Section 4.12, and are highly homologous to ISKra4 family transposases (354). These ISKra4 family transposases are part of the prokaryotic Mutator-like transposases group 4 (p-MULT 4) which differ to other types of transposase as they are flanked by long terminal inverted repeats (TIRs) (355). The ISKra4 is the simplest transposable element in this group, consisting of only one ORF encoding its own transposase (354). Both *sea14* and *sea15* conform to this single ORF encoded transposase construction, as neither ORF has any apparent connection to ORFs surrounding it. In relation to the TIRs, no inverse repeats are present around *sea14* or *sea15*, indicative of degenerate elements, that are probably non-functional.

9.2.2 Se1-3

The region between Sep and Afp on the pADAP plasmid encodes four ORFs homologous to transposases. The first homolog is sea23 which has high homology to IS91 family transposases. IS91 transposases are similar to the IS911 transposases mentioned in section 9.1, in that IS911 was one of the earliest identified transposable elements that could transposition through a rolling-circle mechanism (356) and seems to have a bias for virulence regions (357). In the same region there are three annotations for *se1*, *se2* and *se3* that all seem homologous to IS66 family transposases. Transposases belonging to the IS66 family comprise of three ORFs as is the case in the *se1-3* region (358). They also characterized as containing a 25 bp inverted repeat that is described as being imperfect, in agreement with this, a 37 bp inverted repeat was observed on both sides of the *se1-3* region (Figure 9-2).

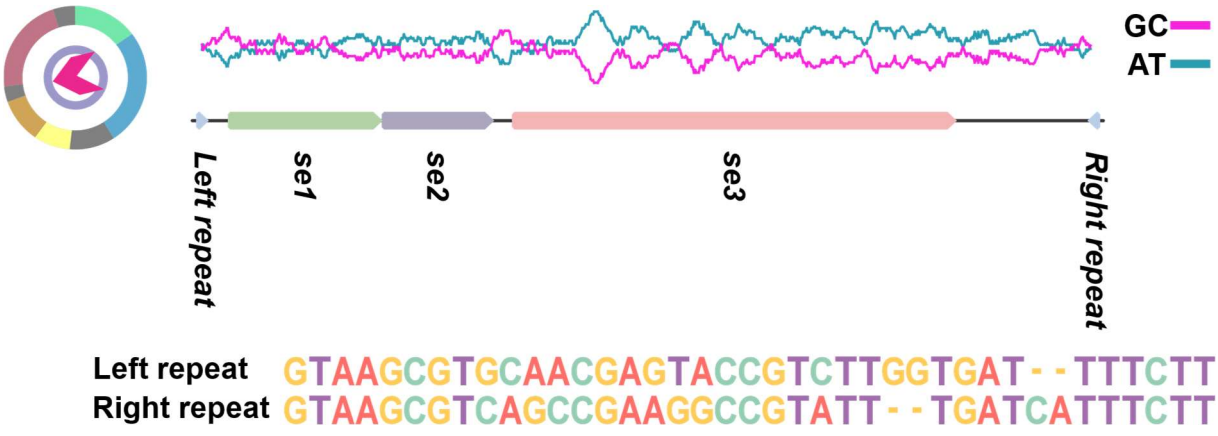


Figure 9-2: The DNA region surrounding *se1-se3*. These three elements are homologous to IS66 family transposases. The two "imperfect" inverted repeats are shown at the bottom and clearly shows a conserved sequence with several mismatches.

9.2.3 Sea24-28

The *sea24* through *sea28* ORFs are found to be highly similar to the ISNCY family of transposases (359). ISNCY stands for IS Not Classified Yet, and therefore no known function has been disseminated yet. The 5' region of *sea24* is identical to the 5' region of *sea27*, and the 3' region of *sea24* is identical to *sea26*. It is therefore assumed that *sea26* and *sea27* were one ORF at one time and *sea24* is most likely a truncated duplication of said ORF. It is very likely that *sea25* inserted into this ORF and is the cause of the truncation of *sea24*, a schematic of the homology between *sea24*, *sea26* and *sea27* can be seen in Figure 9-3. No repeat motifs were identified in the vicinity of this *sea24-sea27* region.

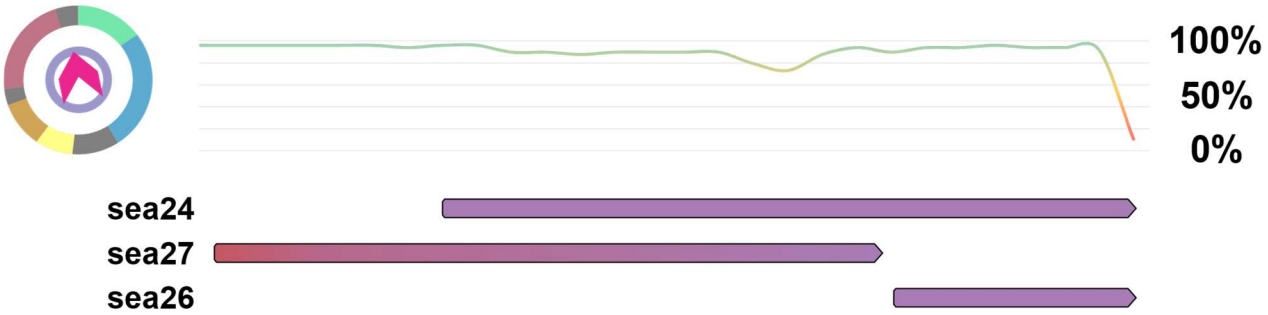


Figure 9-3: The ISNCY family transposase-like ORFs of *sea24*, *sea26* and *sea27* share highly conserved regions. The downstream region of *sea27* is almost identical to the upstream region of *sea24*, the downstream region of *sea24* is almost identical to *sea26*. Potentially *sea26* and *sea27* at some point was a duplication of *sea24* that mutated over time.

9.3 Other potential HGT indicators

Several other possible regions have been identified that may be implicated in HGT, including several hypothetical ORFs, designated *sea5-7*, co-located to the pADAP type VI pilus and ortholog in *Y. ruckeri*. In addition to this, two large DNA repeats, 785 bp in size, have been found which at some point might have played a role in the integration of the PAIs. Oriented in the same direction in pADAP, these repeats were found 44 kb apart, one downstream of the Sep PAI, the other upstream of the type IVa pilus cluster (96, 192).

Both *sea4* and *sea21* share a 785 bp repeat, and it is assumed that these two ORFs are paralogs. They reside on both sides of the *sep*/*sef* region. The pADAP-type plasmid pU143 (83), only has one of these two repeat regions that partially encode *sea4* and *sea21*. Due to the absence of the Afp PAI in pU143, it is postulated that this repeat was obtained by pADAP during the acquisition of the Afp PAI or that the duplication allowed insertion of the island through some unknown mechanism. Based on the data from this study it is unlikely that this repeat element was needed for acquisition of Afp, as only the chronic disease inducing pADAP plasmids, carried by *S. entomophila,* encode for the *sea21* ORF. All other pADAP-type plasmid analysed in this study including the alternate Afp bearing plasmids such as the hyper-virulent AfpX bearing plasmids and the p1137 AfpA bearing variant do not contain the *sea21* element and only encode for *sea4*. Therefore, it is more likely that the shared repeat between *sea4* and *sea21* may reflect a region of low selective pressure that allowed for insertion of large elements without disrupting the functionality of the plasmid.

As shown in this chapter, several elements thought to facilitate active acquisition of novel genetic material, are either different mechanisms than previously thought, such as the *int1* and *in2* ORFs not actually being integrases, were simple transposases that are mostly self-transposing and not affiliated with transposition of entire genetic islands like most IS elements, or similar to other *sea* ORFs, are of a yet to be determined function and have no MGE properties. Based on the assessment of pADAP encoded MGE, this study has found that elements such as *sea4-21* and *int1* are unlikely to actively acquire or excise large GI's from the pADAP plasmid. Further evidence of this was supported by the absence of any observed HGT through *in vitro* and *in vivo* experiments, as outlined in Chapter 10.

Chapter 10 Plasmid stability and fitness experiments

Plasmids are often equipped with a slew of genes that regulate the expression and replication of themselves and the genes contained on them in the cell, and therefore are often seen as a metabolic burden to the cell. This metabolic burden is why plasmids are often considered to be parasitic in nature and are sometimes referred to as selfish genetic elements (SGE). One hypothesis that drove this study was the possibility that plasmids or energy costly genetic elements such as those encoding for toxic effectors would be a handicap to the host cell, and thus active selection to remove them during low nutritional periods would happen so as to alleviate the host from this burden. This hypothesis was broken down in two sub-hypotheses. The first sub-hypothesis was that the plasmid is lost during times of stress by most of the population and that non-plasmid bearing cells would outcompete a small population that maintained the plasmid. The second sub-hypothesis was that active HGT of pathogenicity islands could either facilitate acquisition of novel genetic material or excise pathogenicity islands, and thus lower the metabolic cost of harbouring the plasmid. There are studies, however, that show that co-evolution of plasmid and the primary chromosome can result in lower metabolic cost, thus negating the parasitic traits of the plasmid (156).

10.1 **Pathogenicity islands are not as unstable as predicted**

The production of costly elements such as toxic effectors could be a handicap to the fitness of the host cell, as exemplified by Patel et al. (360) where they found the fitness of non-toxin producing bacteria, they call “cheats”, is higher than those that do produce toxin. Members that do not produce toxic effectors outcompeted the producers (361). Because of the negative effect toxin production supposedly can have on the fitness of the host cell, there is an assumed negative selection for genomic regions that encode for large energy intensive translational products, particularly during periods of low nutrition or stress. To determine if HGT of PAIs or the entire pADAP was happening *in vivo*, the backbones of five pADAP variants, as well as the Sep PAI and Afp PAI were tagged using antibiotic resistant markers. As one hypothesis was that non Afp or non-Sep pADAP-type plasmids could be a result of an excision event as opposed to a precursor of an integration event, the previously published Afp deficient strain, pU143 (83) was also tagged. The final list consisted of six different isolates containing their respective tagged pADAP mutant (Table 10-1).

Table 10-1: Plasmid mutants used to examine HGT of PAIs. pU143 was only tagged in the backbone and Sep region as this plasmid is void of Afp.

Isolate	Plasmid	Backbone	Sep	Afp
145	p145All	Cm^R	Sp^R	Km^R
1100	p1100All	Cm^R	Sp^R	Km^R
210	p210All	Cm^R	Sp^R	Km^R
398	p398All	Cm^R	Sp^R	Km^R
143	pU143BS	Cm^R	Sp^R	-
626	p626All	Cm^R	Sp^R	Km^R

Several other plasmids were tagged, but due to the natural antibiotic resistance of some isolates or lack of recombination, were not carried through to the triple tagged state. The complete list of tagged plasmid mutants can be found Section 2.2.3. Using the plasmid mutants in Table 10-1, two experiments were performed, the first being an *in vitro* culture experiment to see if loss of PAIs happens in optimal lab settings, the second experiment was an *in vivo* bioassay experiment where an attempt was made to observe if PAI loss happens during infection of grass grub larvae.

10.1.1 *In vitro* plasmid and PAI loss experiment

The six plasmid mutants were cultured for 10 days as outlined in Section 2.2.15. No antibiotics were added during the growth. At the final day dilutions were plated out on LB-agar plates devoid of antibiotics. For each plasmid mutant 100 of the overnight grown colonies were patched over to LB-agar plates containing either Chloramphenicol 90 µg per ml (Cm90), Spectinomycin 160 µg per ml (Sp160) or Kanamycin 50 µg per ml (Km50). The Cm90 plates would show presence of plasmid backbone, the Sp160 plates the presence of the Sep PAI and the Km50 the presence of the Afp PAI. Of the 100 assessed colonies for each plasmid mutant, growth on all antibiotic plates was observed. This leads to the assumption that no loss had happened in these optimal growth conditions or at least it happened in a rate lower than 1:100.

10.1.2 *In vivo* plasmid and PAI loss experiment

Another experiment undertaken with the six plasmid mutants was performing a bioassay as outlined in Section 2.4.1. *Serratia* bacteria were extracted from the grub after 12 days as outlined in Section 2.3.1. For each plasmid mutant 100 of the overnight grown colonies were patched over to LB-agar plates containing either Cm90, Sp160 and Km50 and assessed as described in Section 10.1.1. The observation was made that growth happened on all antibiotic plates. Therefore, it can be concluded that no excision of PAIs had happened *in vivo*, or that it happens in a rate lower than 1:100.

10.2 **Plate experiments to confirm plasmid loss**

Eleven representative strains for each of the pADAP genotype were analysed for the stability of the respective plasmid they were carrying. The plasmids in isolates 149, 299, 465, 626, 1048, 1129, 1137, 1769, D, Agr96X and Puna18, were tagged with a Cm^R tag, as explained in Section 2.2.14. To determine the stability of the pADAP-type plasmid in their respective hosts, these strains were cultured in different growth conditions to be further assessed at the ten-day time point. The growth conditions were either regular LB-broth, M9 minimal medium broth with 0.4% casamino acid as carbon source or LB-broth spiked with 0.5 µg / ml of Mitomycin C (MitC). The M9 growth medium would simulate conditions of nutritional stress, known to facilitate loss of larger plasmids (362, 363). The MitC can induce DNA crosslinking, a lethal event for bacteria, and this action activates SOS-responses in the host (364-366). SOS-response in bacteria is tied to loss of plasmid, through halting replication, preventing segregation/conjugation or destabilizing the plasmid through direct MitC induced DNA damage, although not every plasmid is affected equally (367, 368). The experiment was performed as described in Section 2.2.15. The final Colony-forming units (CFU) were grown on plates containing Cm90 and were scored for growth. All assessed colonies appeared to still be Cm90 resistant as shown in Table 10-2 and therefore conclude that no plasmid loss occurred or happened in a ratio lower than 1:200. The summary data for this experiment can be found in Supplementary Materials S.32.

Table 10-2: Plasmid retention after 10 days for all 11 representative pADAP genotypes. Table contains the results of an antibiotic plate screening of colonies obtained after 10 days on different culture conditions. Dilutions of the cultures were plated out on LB, and 200 colonies grown on those were patched over to Cm90 LB-agar plates. If colonies grew on the patched plates, it was assumed the plasmid was retained as non-tagged wildtype strains did not grow on Cm90 LB-agar plates.

Isolate	LB	M9	MitC
149	200/200	200/200	200/200
299	200/200	200/200	200/200
465	200/200	200/200	200/200
626	200/200	200/200	200/200
1048	200/200	200/200	200/200
1129	200/200	200/200	200/200[1]
1137	200/200	200/200	200/200
1769	200/200	200/200	200/200
D	200/200	200/200	200/200
AGR96X	200/200	200/200	200/200[2]
Puna18	200/200	200/200	200/200
Average	100.0%	100.0%	100.0%

[1] The culture density was not stable and dropped to 10^6 CFU /ml on two days but recovered the day after.

[2] A formation of a ring of lytic cell debris could be observed on certain days on top of the MitC culture.

10.3 **Flow cytometry experiments to confirm plasmid loss**

The experiment described in Section 10.2 was repeated using a different method of analysing plasmid loss per cell. The plasmids in isolates 149, 299, 465, 626, 1048, 1129, 1137, 1769, D, Agr96X and Puna18, were tagged as explained in Section 2.2.14 with a Green Fluorescence Protein mutation 3 (GFPmut3) (213) tag from the pUC30T_GFPmut3 plasmid described by Barbier et al (212). The isolates containing the GFPmut3 were cultured as described in Section 2.2.15. At the tenth day the samples were prepared for flow cytometry as described in Section 2.2.16. During the experiment there appeared to be a noticeable number of events that were not actual cells passing through the capillaries, since most prokaryotic cells are significantly smaller than the eukaryotic cells for which these machines are often intended. This resulted in a small amount of error in the analysis, but nevertheless similar results to those found in Section 10.2 was observed. In Table 10-3 it can be seen that the negative control did not contain any fluorescent cells on the Fluorescein isothiocyanate (FITC) channel, meaning the cells do not get excited at ~495 nm or do not emit at ~519 nm, so all registered fluorescent events are true events. The positive control reveals the previously mentioned error of non-cell events registered as the events measured are not all fluorescing. The isolates analysed all have a remarkable 85% or higher retention of plasmid based on analysis of thousands of cells, even in harsh conditions such as the MitC induction. Therefore, it is highly unlikely that cells can easily lose the pADAP plasmid in short time periods, a longer lasting study would be able to confirm this further.

Table 10-3: Plasmid retention after 10 days for all 11 representative pADAP genotypes. Table contains the average percentage of fluorescent events in the gates that were predetermined based on the positive and negative controls. Elimination of all background noise was unsuccessful, resulting in 2,1% of the positive sample still reporting non-fluorescent events, meaning similar results are to be expected in the other samples.

Isolate	LB	M9	MitC
149	94.3%	92.5%[1]	86.2%
299	92.8%	85.6%	83.5%
465	93.2%	90.0%	88.1%
626	93.4%	89.8%	91.8%
1048	87.6%	90.1%	89.9%
1129	93.1%	94.9%[2]	85.0%
1137	91.4%	91.4%	93.7%[1]
1769	94.4%	90.4%	90.0%
D	90.8%	90.8%	86.2%
AGR96X	91.9%	89.3%	94.0%
Puna18	89.6%	89.0%	87.1%
Average	92.1%	90.3%	88.7%

Controls	LB
Negative	0%
Positive	97.9%

[1] One replicate in these samples displayed events that did not conform to the predefined gates, therefore the replicate was removed.

[2] One replicate could not be analysed due to a break in the vial, preventing the machine from forming a vacuum.

10.4 **Fitness cost of pADAP**

As shown in the previous section of this chapter, there is strong evidence that the pADAP plasmids are not easily lost, and act more as a secondary chromosome or chromid. The question then became, why is plasmid loss not happening? To find this out a 24 h fitness test between A1MO2 wildtype strain from which pADAP was first isolated (218) and 5.6 which is a pADAP heat cured mutant (218) was performed. The experiment was run as described in Section 2.2.17. The results of the experiment are shown in Figure 10-1.

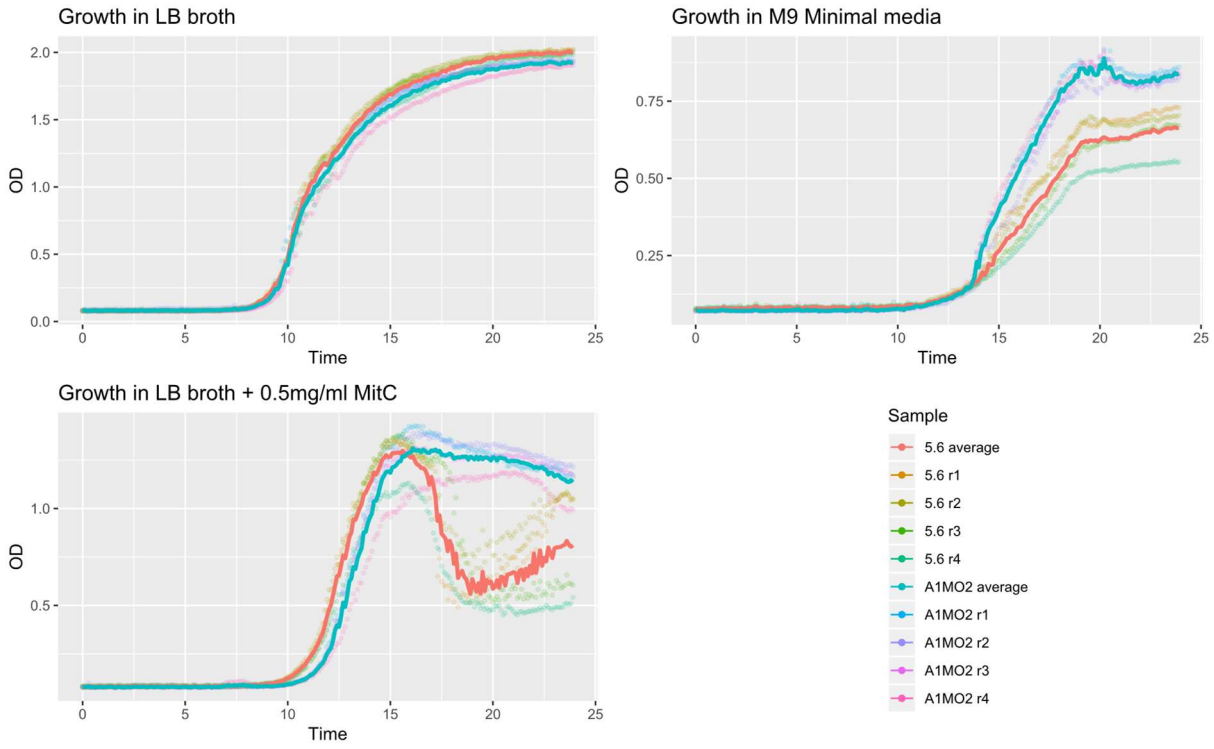


Figure 10-1: Several growth experiments performed on a plasmid bearing wildtype and its plasmid null counterpart. The 5.6 mutant is a pADAP heat-cured mutant of A1MO2 typestrain (81).

From this experiment no difference in fitness between plasmid bearing and plasmid null cultures were found under optimal laboratory growth conditions. There is no real observable difference between the A1MO2 and 5.6 replicates. However, under minimal medium and MitC growth conditions, there is a strong fitness benefit for the pADAP carrying A1MO2 as opposed to the plasmid null 5.6. The drop of 5.6 growth after reaching the highest cell density point is of note. These *Serratia* bacteria harbour numerous bacteriophages (369) which are known to be produced when SOS response is triggered, which is something MitC does.

The pADAP plasmid contains two SOS-repressor genes *psiA* (370) and *psiB* (371). It is most likely that the presence of *psiA* and *psiB* limit the number of cells entering the lytic cycle, which in turn limits production of the chromosomal bacteriophages under stress condition, which it in turn limit the rate cells in the population being affected with lysis. Further work needs to be undertaken to confirm that this is the actual reason behind the fitness benefits of pADAP under MitC induction. The slightly higher stationary phase CFU under M9 growth conditions could be a result of pADAP suppressing transcription of chromosomal genes as some plasmids are known to do (127), most notably genes correlated to metabolism (129), thus conserve energy for reproduction, resulting in a higher cell density. No such experiment has been performed using RNAseq on pADAP to confirm this.

Chapter 11 Toxin Anti-toxin

Due to the assumed parasitic nature of plasmids it is often believed that the host would lose the plasmid over time due to its sometimes-unfavourable selection traits. However, many plasmids are stable, even in stressful conditions as exemplified in Chapter 10. The duality of the plasmid being a burden that should be negatively selected on, but meanwhile are found wide spread and are hard to remove even in optimal lab settings is referred to as the 'plasmid paradox' (154).

One key feature of parasitosis is the inability of a host to easily remove the parasite (372). Plasmids can employ several ways of forcing their perpetuation inside a bacterial population. One of these systems is referred to as PSKS, sometimes also called addiction systems (373) which involves TA genes (143). Toxins are accumulated in the cytoplasm and during segregation will be distributed among daughter cells. To survive, these cells will require an, often less stable, RNA or protein-based anti-toxin to counteract the toxin. Another type of TA are systems that secrete anti-microbials, and only anti-toxin carrying cell can survive in their surroundings. Variations of these two exist. Several pADAP-type plasmids in this study were found to have either one or several TAs or TA-like systems present, which could explain their persistence in the population.

11.1 HigAB

The translated products of the ORFs sea40 and sea41, located in the backbone of all pADAP-type plasmids, share AA identity to HigAB TA orthologs (Table 11-1). This TA operon is present in the conjugation region of the backbone as previously described in Section 4.8.1 and shown in Figure 11-1.

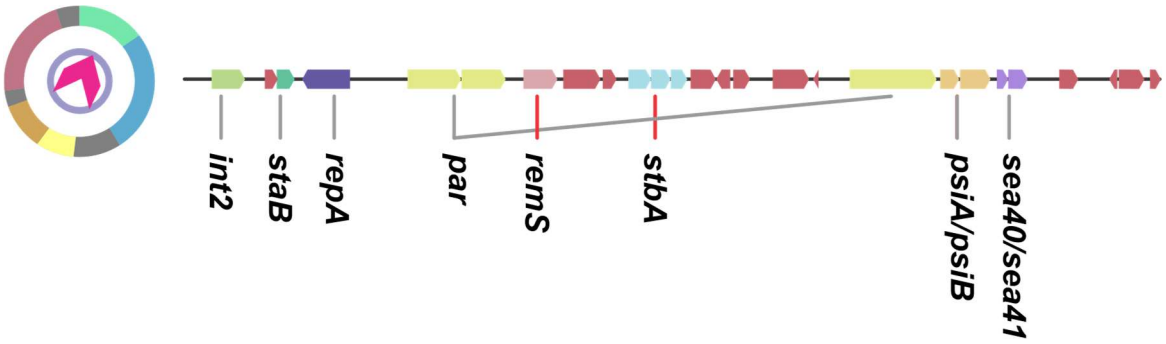


Figure 11-1: Replication region of the pADAP backbone. The HigAB orthologous *sea40*/*sea41* operon is situated downstream to the *psiA*/*psiB* SOS-repressor operon.

This group of HigAB TAs are bacteriostatic, meaning they do not destroy the cell, but instead induces growth arrest. This group is involved with cleavage of mRNA (374) and thus inhibiting translation of certain genes similar to the VapBC system of TAs (375). The HigB (Sea40) protein is the toxin in this ORF cluster and the HigA (Sea41) the antitoxin that negates the functionality of HigB through docking with HigB.

An orthologous TA found in *Salmonella choleraesuis* serotype typhimurium, designated SehAB was shown to autorepress its own operon and can disrupt the autorepression in periods of nutritional stress (376). The paper by De la Cruz et al. (376) on SehAB demonstrates the importance of the TA in the ability of *S.* typhimuriums to colonize mice lymph nodes.

The HigAB TA system is also associated with biofilm formation and production of virulence factors in *Pseudomonas aeruginosa* (377) and *higB* (under the ORF name *ympt1.66c)* has also been determined to be one of the key factors in the ability of *Yersinia pestis* to infect a mammalian host with the bubonic plaque by enabling the bacteria to persist in macrophages (378).

Unfortunately, this link between HigAB and the Sea40 and Sea41 was not identified until the end of the PhD and therefore no experiments have been performed to show the correlation between this cluster and colonization, virulence or plasmid stability.

Table 11-1: BlastX hits of sea40 and sea41. These elements seem to encode a HigAB orthologous TA.

Protein	Hit						Accession	
Sea40	cytoplasmic protein [*Serratia grimesii*]						WP_061807487.1	
	Length:	103	Range:	3 to 103	Eval:	8E-68		
	Ident:	99/101(98%)	Pos:	100/101(99%)	Gap:	0/101(0%)	Coverage:	98.0%
	mRNA interferase HigB (modular protein) [*Serratia symbiotica*]						CDS57329.1	
	Length:	157	Range:	57 to 157	Eval:	8E-65		
	Ident:	96/101(95%)	Pos:	99/101(98%)	Gap:	0/101(0%)	Coverage:	95.1%
Sea41	MULTISPECIES: helix-turn-helix domain-containing protein [*Serratia*]						WP_127147056.1	
	Length:	147	Range:	1 to 147	Eval:	4E-101		
	Ident:	144/147(98%)	Pos:	147/147(100%)	Gap:	0/147(0%)	Coverage:	98.0%
	type II toxin-antitoxin system HigA family antitoxin [*Salmonella enterica*]						EAB5051308.1	
	Length:	148	Range:	1 to 147	Eval:	4E-62		
	Ident:	93/147(63%)	Pos:	117/147(79%)	Gap:	0/147(0%)	Coverage:	63.3%

Table 11-2: BlastX hits of the two TA ORFs found on p1129. These elements seem to encode a Colicin orthologous TA.

Protein	Hit						Accession	
TA OrfA	lipid II-degrading bacteriocin [*Pantoea ananatis*]						WP_110286972.1	
	Length:	270	Range:	5 to 267	Eval:	9E-73		
	Ident:	118/263(45%)	Pos:	170/263(64%)	Gap:	5/263(1%)	Coverage:	44.9%
	colicin M-like protein [*Pantoea* sp. SJZ147]						TWD38167.1	
	Length:	133	Range:	1 to 130	Eval:	2E-41		
	Ident:	69/130(53%)	Pos:	92/130(70%)	Gap:	1/130(0%)	Coverage:	53.1%
TA OrfB	colicin M resistance protein [*Pantoea* sp. Sc1]						EIB96703.1	
	Length:	122	Range:	11 to 122	Eval:	3E-14		
	Ident:	38/116(33%)	Pos:	65/116(56%)	Gap:	4/116(3%)	Coverage:	32.8%
	Microcin-M immunity protein [*Enterobacter cloacae*]						VAF17058.1	
	Length:	117	Range:	1 to 115	Eval:	2.2		
	Ident:	31/125(25%)	Pos:	61/125(48%)	Gap:	10/125(8%)	Coverage:	24.8%

11.2 **Colicins**

Five pADAP-type plasmids were found to encode for a colicin orthologous TA. Colicins are secreted proteinaceous anti-microbial compounds (146). These colicins are frequently paired with a plasmid encoded immunity proteins (379), to form a TA system.

There are many types of colicins, some associated with pore-forming, some with DNAse or RNAse activity, and one that is associated with the degradation of cell membrane lipids (380). Colicins exploit different outer membrane protein (OMP) transport systems, mainly the Tol-Pal system (Type A) and the TonB transport system (Type B), to get internalized (381).

Colicin systems are commonly found on *E. coli* associated plasmids, but have been found in the Yersiniaceae family of bacteria as well, such as a Type A colicin 28b homolog in *Serratia marcescens* (382). The plasmids identified in this study which encode a colicin TA encode for a Type B Colicin M homolog lipid II-degrading bacteriocin (Table 11-2) (383). These ORFs appear to be the first colicin M genes identified the *Serratia* genus. Colicin M is transported across the OM through a FhuA membrane transporter (384). This FhuA protein has a cork domain that “opens” upon interaction with the TonB box domain (385). The TonB box domain facilitates the transition of ColM through the FhuA transporter. Once the ColM enters the periplasm, it degrades the peptidoglycan structure (386). The process with which ColM disrupts bacterial cell activity is shown in Figure 11-2.

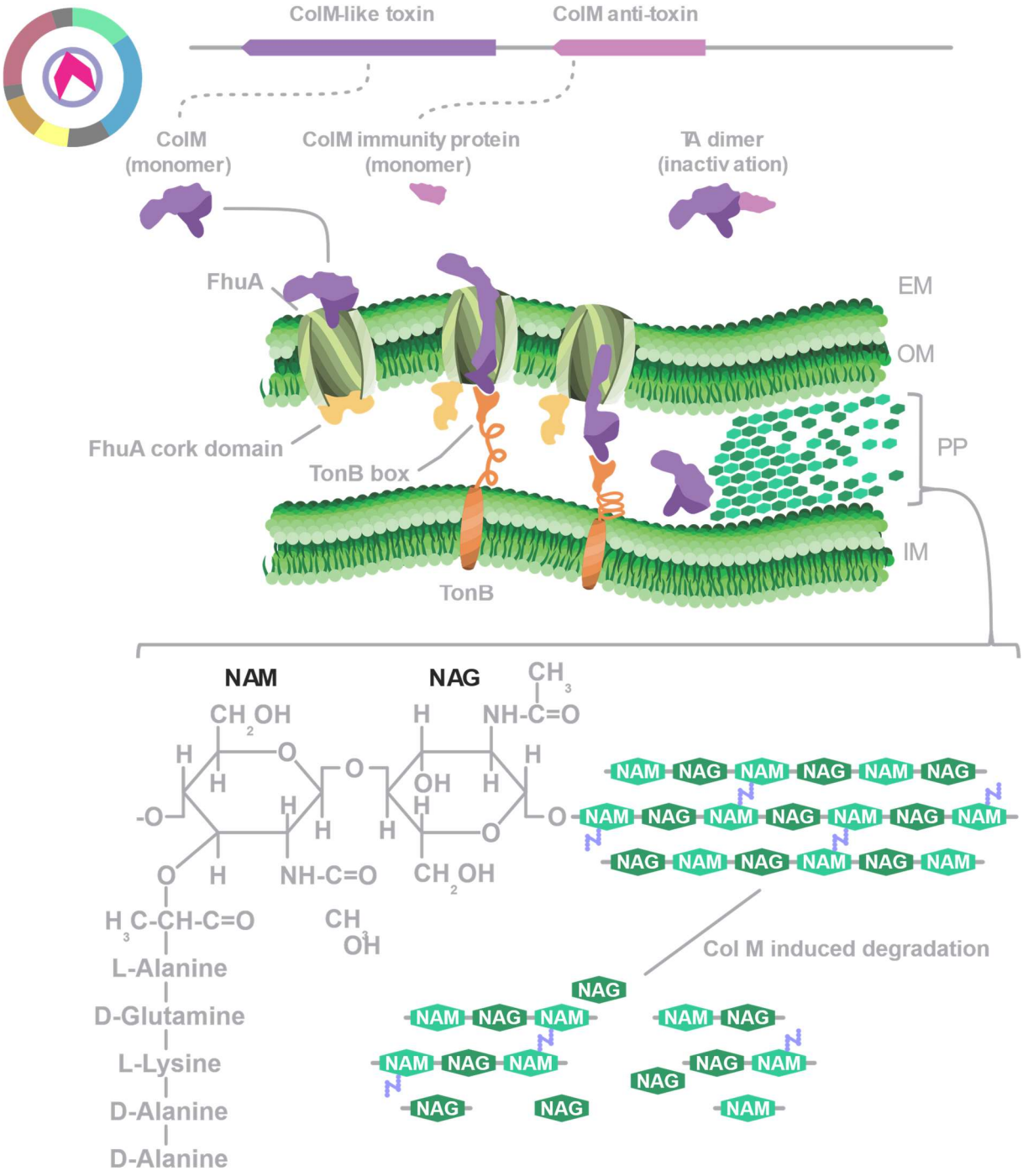


Figure 11-2: An illustration showing the process with which Colicin M disrupts bacterial cell activity. Interaction of ColM with FhuA and the TonB box, results in internalization of ColM into the periplasm. Here the ColM degrades the peptidoglycans resulting in destabilization of the IM.

The region on which the colicin TA operon is encoded is positioned 3’ of the Afp encoding cluster, as characterized in Section 8.3. The colicin M region can be seen in Figure 8-4, and has only been identified on pADAP-type plasmids, p4, p163, p1129, p1A, pSpF.

Often only a subpopulation of the actual colicin encoding cells are producing the toxin as a direct effect of the arms race for nutrients between sensitive cells, resistant cells and toxin producing cells with encoded anti-toxins (145).

Manual assessment of chromosome annotations of isolate 1129 revealed no additional chromosomally encoded bacteriocins. To confirm that bacteriocins were actually produced by the colicin orthologous TA operon, the colicin region was deleted using the methods described in Section 2.2.11, Section 2.2.12 and Section 2.2.13, creating the p1129Δcolicin plasmid mutant (Section 2.2.3).

To determine the anti-microbial effects of this colicin M encoding region against other bacteria, the 1129 WT and 1129 containing the p1129Δcolicin were cultured o/n, and bacteriocin extracted using the purification method described by Hockett et al. (224), as outlined in Section 2.2.18. The protocol by Hockett et al. was then used to perform a soft-agar overlay screening using both the WT and the mutant to determine presence of bacteriocins. The indicator bacterium used in the soft-agar overlay was Dh10B (215). A 5 µl droplet of a 1:1, 1:5 and 1:10 dilution for each sample was added to the soft-agar overlay. Plates were left for approximately 6 h, and clearing was only observed in droplets from the WT sample, but regular growth was observed in the p1129Δcolicin mutant bearing isolate (Figure 11-3). This indicates that a bactericidal bacteriocin was produced by the TA operon on the p1129 plasmid.

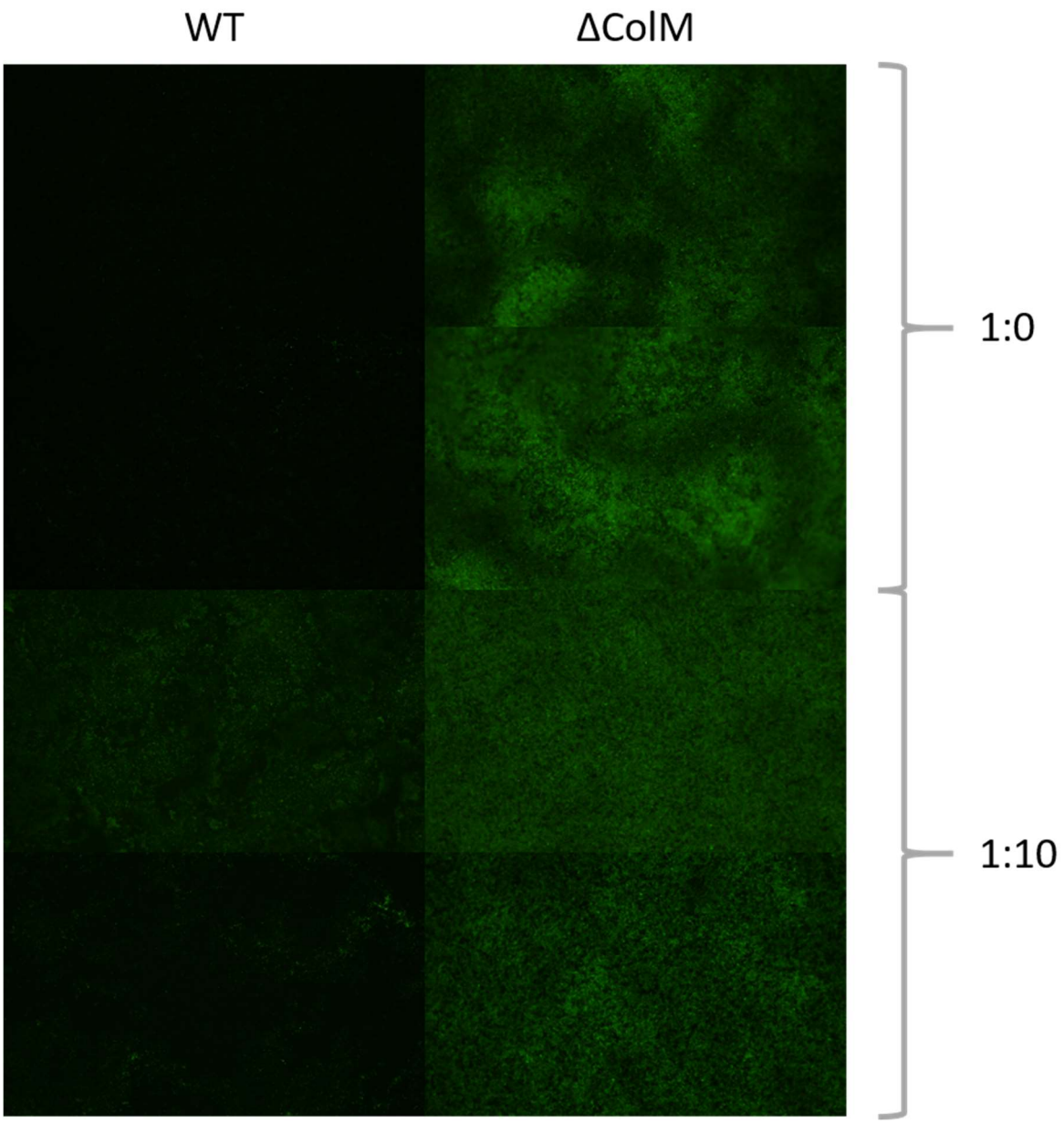


Figure 11-3: 100x microscope pictures of plates spotted with purified protein samples. The left column represents pictures taken from ColM containing spots, purified from WT isolates. Right column represents spots from purified ΔColM mutant sample. Indicator strain was Dh10B + pUC30TGFPMut3

These bacteriocins act as anti-microbials in nature, reducing competitive bacteria, but also act as PSKS by destroying plasmid free segregants.

11.3 **VapBC**

Four pADAP-like plasmids as well as one non-pADAP-type plasmid in this study were found to contain a VapC homolog toxin and an AbrB homolog antitoxin, containing a VapB domain (Table 11-3, Table 11-4).

Table 11-3: Isolates containing a VapBC orthologous TA encoding plasmid.

Species		Isolate
pADAP-type		
Serratia	*entomophila*	AGR96X, Sprot5, LC, 20093
Non-pADAP		
Serratia	*proteamaculans*	220

The VapBC locus, found downstream of several AfpX encoding pADAP-type plasmids as previously described in Section 8.3, encodes a bacteriostatic TA system, similar to HigAB mentioned in Section 11.1. The VapBC loci is one of the most common Type II TAs, Type II referring to the mechanism to nullify the toxin is by a protein based antitoxin, interacting directly with the protein based toxin (147).

The antitoxin can negatively autoregulate the *vapBC* operon by binding to the toxin and forming an octamer, consisting of four VapB and four VapC particles (387) (Figure 11-4). This $(VapB_2C_2)_2$ octamer structure is sometimes referred to as the VapBC octamer (VapBCO) complex (388). The VapBCO complex is formed when a VapC homodimer binds to the VapB homodimer and form a “bridge”. Two VapBC “bridges” are required for the formation of the VapBCO complex (389). Abundance of VapC monomers can destabilize the VapBCO complex by invading the VapBCO complex. This results in the VapBCO complex associated VapC to form dimers with the free-floating VapC monomers and breaking away. This in turn causes the release of the DNA bound VapBCO complex and leads to resumption of transcription of the operon. This concentration dependant regulatory system is also referred to as conditional-cooperativity (388). The operon of TA loci most commonly starts with the antitoxin gene such as *vapB*.

Table 11-4: BlastX hits of the two TA ORFs (*orfA* and *orfB*), located downstream of the AfpX region.

Protein	Hit						Accession	
TA *orfA*	type II toxin-antitoxin system VapC family toxin [*Brenneria roseae*]						WP_136157135.1	
	Length:	138	Range:	1 to 138	Eval:	8E-92		
	Ident:	131/138(95%)	Pos:	135/138(97%)	Gap:	0/138(0%)	Coverage:	94.93%
	PIN domain-containing protein [*Yersinia massiliensis*]						WP_145494113.1	
	Length:	138	Range:	1 to 138	Eval:	7E-90		
	Ident:	126/138(91%)	Pos:	133/138(96%)	Gap:	0/138(0%)	Coverage:	91.30%
TA *orfB*	AbrB/MazE/SpoVT family DNA-binding domain-containing protein [*Kosakonia* sp. WCHEs120001]						WP_131413711.1	
	Length:	76	Range:	1 to 76	Eval:	5E-44		
	Ident:	71/76(93%)	Pos:	74/76(97%)	Gap:	0/76(0%)	Coverage:	93.42%
	Virulence-associated protein vagC [*Klebsiella pneumoniae*]						AWM64240.1	
	Length:	89	Range:	14 to 89	Eval:	4E-43		
	Ident:	69/76(91%)	Pos:	73/76(96%)	Gap:	0/76(0%)	Coverage:	90.79%

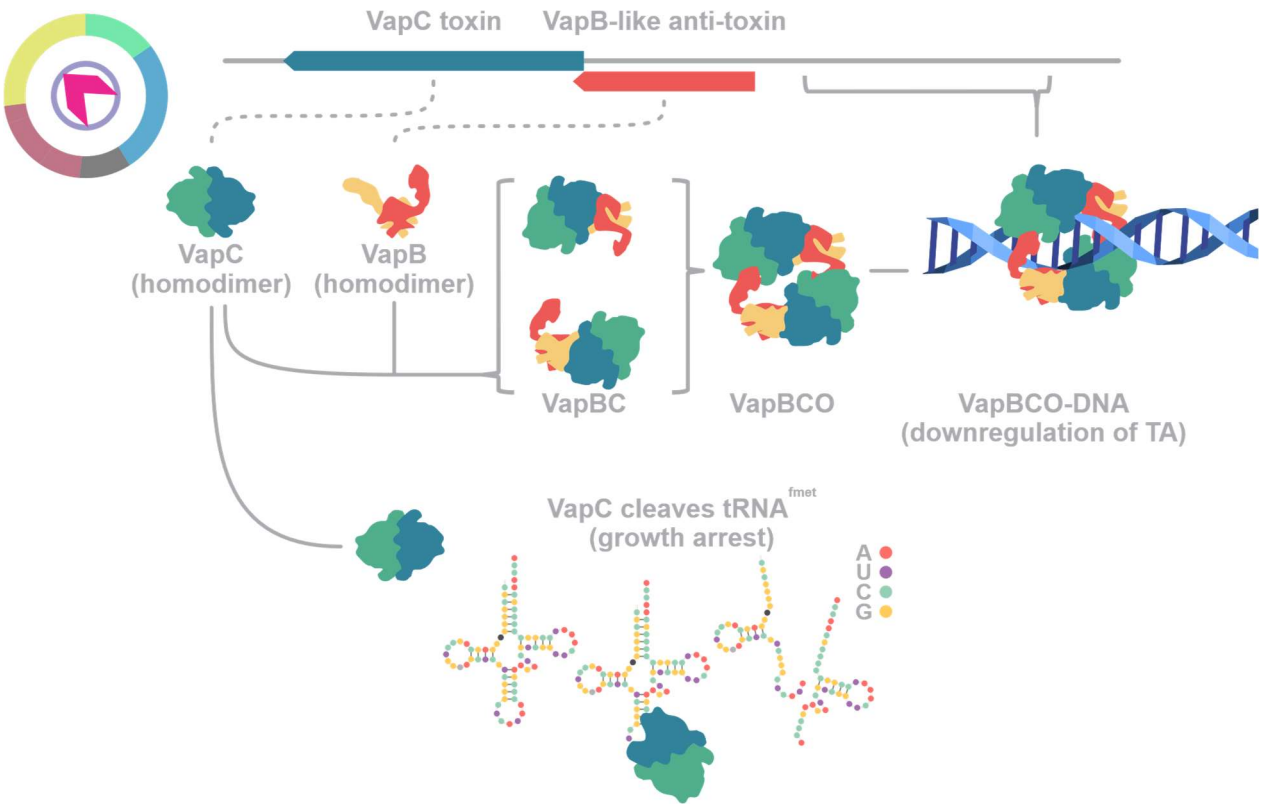


Figure 11-4: An illustration of the VapBC complex modes of function. If equal amounts of VapB and VapC are present, two homodimers of VapB and VapC can form the VapBC complex. Two VapBC complexes can then in turn form the VapBCO complex that is able to upregulate the transcription of the VapBC operon. If VapC is present in a higher concentration than VapB, it is able to bind to $tRNA^{fMet}$ and cleave it, causing growth arrests by preventing global translation of mRNA.

VapC is known for its ability to cleave the prokaryotic initiator Transfer RNA N-Formylmethionine ($tRNA^{fMet}$) (390). The tRNAs are required for interfacing between messenger RNA (mRNA) and AAs, and the synthesis of proteins exclusively starts with a methionine derivative containing a formyl group, making $tRNA^{fMet}$ essential for translation of all prokaryotic mRNA. The VapC cleaves the $tRNA^{fMet}$ between the anticodon stem and loop, preventing the tRNA from being used to initiate translation, thus depleting the cell of its initiation tRNA, and preventing global cellular translation (375). Cleavage is performed by a ribonuclease PilT N-terminal (PIN) domain (391).

During nutritional stress the unstable VapB antitoxin is degraded more easily by ATP-dependant Lon proteases, making more VapC monomers available to deplete $tRNA^{fMet}$ (388). This results in a halted cellular translation, including that of the *vapBC* operon, eventually spiralling into a complete halt of cellular translation due to the lack of freely available $tRNA^{fMet}$. This leaves the cell in a metabolic persister state, where it is metabolically dormant until exposed to new nutrients (392).

As a side effect of the dormant state, the cell becomes more tolerant to xenobiotic effectors that disrupt metabolic pathways (148) and could potentially increase shelf life of the biopesticide product, or could be used to re-activate dormant cells a season after application.

To determine if the plasmid is less stable with the absence of the VapBC complex *in vitro*, the 10 day growth experiment described in Section 10.1.1 was performed with the presence of the XΔNOVa VapBC deficient mutant Hurst [unpublished data] (Section 2.2.3).

The results (Table 11-5) reveal that the plasmid is stable within the confines of the experiment regardless of absence or presence of the VapBC construct. The table contains the results of an antibiotic plate screening of colonies obtained after 10 days on different culture conditions. Dilutions of the cultures were plated out on LB and patched overnight to LB plates containing Cm for AGR96X and Spec for XΔNOVa respectively to determine plasmid loss.

Table 11-5: Plasmid retention after 10 days for all AGR96X and the VapBC null XΔNOVa mutant.

	LB	M9	MitC
AGR96X	100/100	94/94[2]	100/100[1]
XΔNOVa	100/100	100/100	100/100[1]

[1] Lytic-rings were seen on top of the broth at several time points.

[2] Not enough colonies were present to assess 100.

Chapter 12 Evolutionary divergence

The previous chapters described the relationship between plasmid sequences obtained from the *S. entomophila*, *S. ficaria*, *S. grimesii, S. liquefaciens*, *S. marcescens*, *S. proteamaculans*, and *Y. frederiksenii* isolates analysed in this study. This chapter is intended to provide a comprehensive summary of the diversity within the pADAP plasmid family.

12.1 **Overview of pADAP-types**

The backbone of all pADAP-type plasmids are relatively conserved (Chapter 4) albeit several insertions have been observed (Section 4.7, Section 4.8). The most notable insertion in the backbone is a ~3.9 Kb insertion between the *traG* and *trbC* gene (Section 4.9). This 3.9Kb inserted region is only present in pADAP-type plasmids harbored by *S. proteamaculans* and *S. liquefaciens* isolates and therefore is a good indicator of an evolutionary split between pADAP and the other pADAP-type plasmids. Apart from smaller regions of divergence between the member of the pADAP family plasmid backbones (as defined in Sections 4.7 through 4.8.3), several highly divergent backbones were identified (Figure 12-1).

The *S. proteamaculans* pADAP-type plasmid p465 encodes a putative functional bacteriophage, designated as Spb, inserted between the Type IVa pili ORFs *pilO* and *pilN*. The ~35 Kb Spb encoding region shares 88% DNA similarity with an homologous bacteriophage encoding region of the *S. proteamaculans* non-pADAP plasmid p591 (Section 4.10, Section 8.4).

The *S. proteamaculans* p1769 pADAP-type plasmid is missing approximately 23 Kb of the conjugative region from *traT* to *pilL*, as outlined in Section 4.10, but surprisingly still encodes a variant of the demarcation region as discussed in Section 4.12. As described in Section 12.4.2, the 1769 isolate appears to not be affected in its ability to conjugate the plasmid to other *Serratia* cells.

The last highly divergent pADAP-like plasmid is that of *S. marcescens* WVU-005-1, sourced from the NCBI RefSeq database, originating from a US based clinical sample. Though encoding a different replication region as outlined in Section 4.11 and Section 4.12, the backbone region of WVU-005-1 has a surprisingly high DNA similarity of 79% to pADAP (Figure 4-7). This is of interest as the plasmid is not associated with the endemic grass grub.

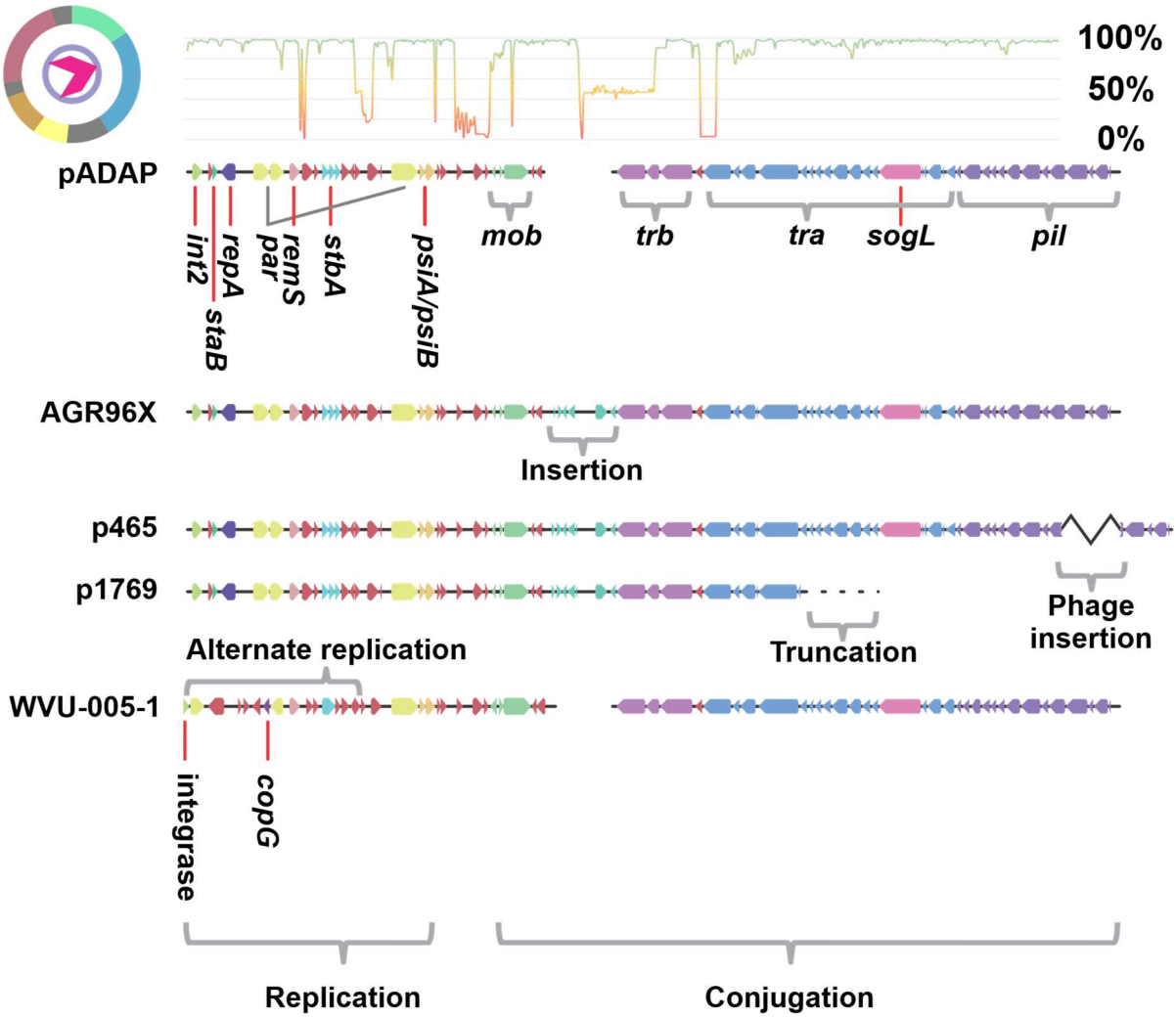


Figure 12-1: The five most notable divergent pADAP-backbones. The AGR96X variant is shared with numerous other S. proteamaculans and S. liquefaciens derived pADAP-type plasmids.

Due to the highly conserved gene synteny, shared throughout all these plasmids including the WVU-005-1, these pADAP-type plasmid backbones, starting from *int2* to *pilL* can therefore be considered a *Serratia* associated capturing apparatus (SACA) that seems specific to *Serratia* spp. The core genetic make-up of the SACA will be discussed in Section 12.2.

The elements captured by this device include the previously published Sep (Chapter 5) (161) and Afp (Chapter 7) (161) encoding PAIs, including variants of the Afp such as the AfpX PAI (69), which appear highly stable as outlined in Chapter 10. In this study several novel regions have been identified such as the RUF GI (Section 8.1), several conserved regions downstream of Afp and it's variants (Section 8.3), a cluster of ORFs found on WVU-005-1 (Section 8.6) and a predicted plant associated GI found on pPuna18 (Section 8.4). These data are summarized in Figure 12-2, showing the synteny map for 12

different pADAP-like genotypes. The isolates in which each pADAP-like genotype was identified are listed Table 12-1.

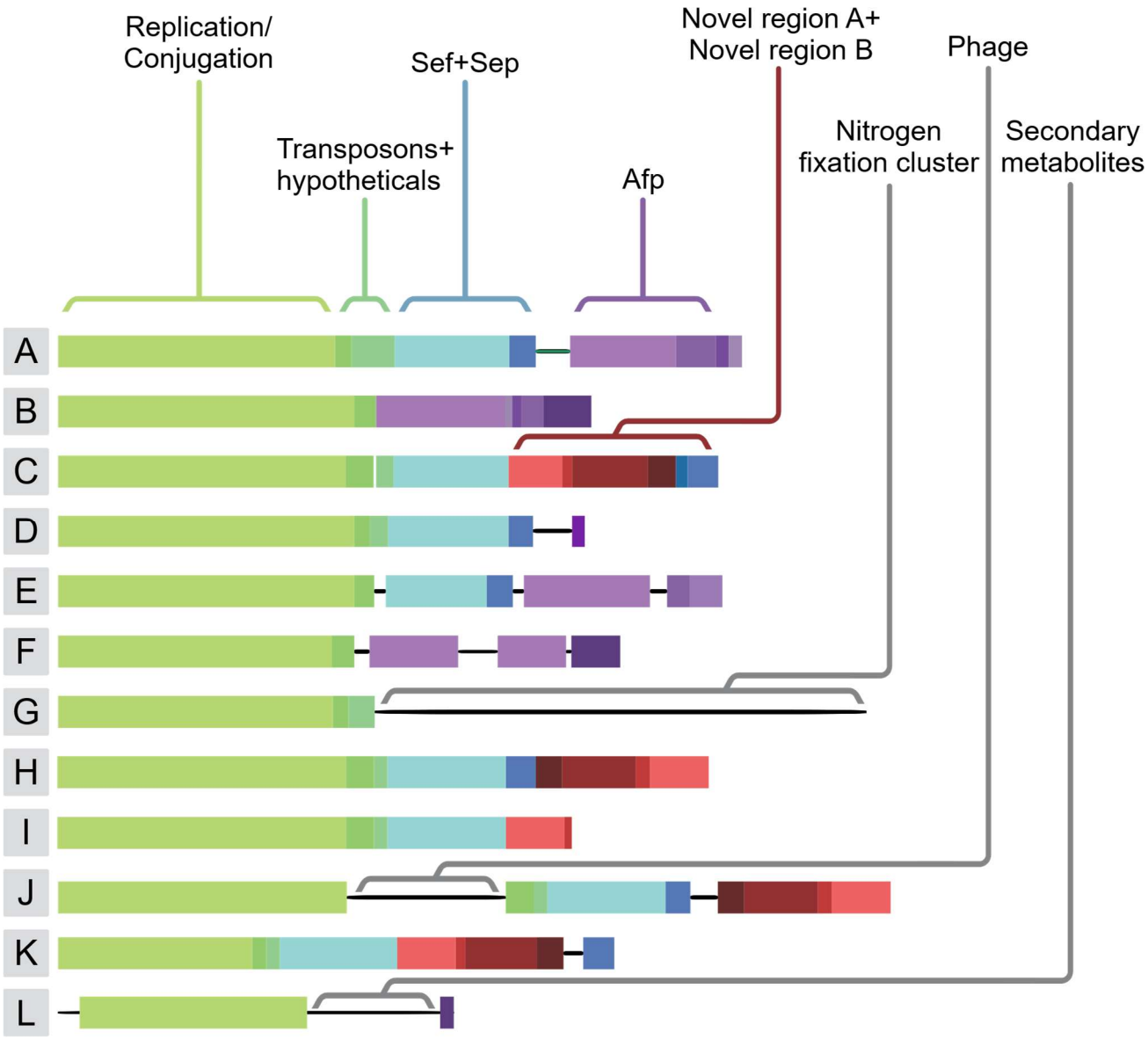


Figure 12-2: Synteny map of the pADAP family plasmid types. The reference pADAP plasmid [NC_002523] (83) is depicted as type A, the plasmid obtained from strain AGR96X (69) is depicted as type B. Colours symbolizes regions of interest as outlined in previous chapters and sections. Green the backbone region (Chapter 4), Blue the Sep and Sef region (Chapter 5, Chapter 6), purple the Afp region (Chapter 7), red the novel RUF region (Chapter 8) and unmarked regions are unique to the plasmid type in question and described in the figure. Synteny blocks are based on progressiveMauve alignment of all pADAP-type plasmids (393).

Table 12-1: pADAP-like plasmid genotype per isolate.

Genotype[1]	*Serratia* species	Isolate
A	*entomophila*	158, 176, 210, 345, 398, 626, 1100, Diarr, Moraki_2, A1MO2
	proteamaculans	142, 145
B	*proteamaculans*	4, 163, 1A, 20093, SpF, LC, AGR96X, Sprot5
C	*proteamaculans*	336, 1770, D
D	*liquefaciens*	376, 377
	proteamaculans	1, 143, 1048, 1071, 10novel, 12newD, 25E, CfB, E, G, R10, RM5
E	*proteamaculans*	1129
F	*proteamaculans*	1137
G	*proteamaculans*	Puna18
H	*proteamaculans*	149, 1772, 12a, 12d, 28F, M
I	*proteamaculans*	299, 1457, K
J	*proteamaculans*	465
K	*proteamaculans*	1769
L	*marcescens*	WVU-005

[1] Genotype defined in Figure 12-2

With reference to Figure 12-2 and Table 12-1, plasmids belonging to group C and K contain both the Sef (light blue), Spp (dark blue) and RUF (red) GIs, however these plasmids have the Sep/Spp and RUF region inversed relative to the orientation of Sep and RUF found in most pADAP-type plasmids, i.e. the Spp/RUF region is not 5' > 3' downstream of Sef, but instead is 3'>5' (Figure 12-2). It is not clear whether this rearrangement of the RUF GIs orientation is due to independent acquisition or due to genomic reshuffling. These plasmids, except for pD, have a degenerate SppA directly downstream of the Sef island shown as 'Degenerate A' in Section 5.2 and Figure 5-2. It is therefore likely that in plasmids p336, p1770, p1769 and potentially pD, the Spp has undergone a rearrangement event at some point in time. Based on coverage plots, there appears to be equal amounts of plasmid DNA to chromosomal DNA (Section 3.5), strongly suggesting these plasmids are single copy, and no isolates were identified to carry more than one pADAP-type plasmid, most likely due to a process called handcuffing (Section 4.2).

12.2 **Genetic diversity within genotypes**

Although 12 distinct pADAP-type plasmids were identified (designated A-L in Figure 12-2), there is still significant genetic diversity within each group. As outlined in Section 3.4, plasmids were annotated using the Prokka feature annotation tool (249). The CDSs of all pADAP-type plasmid, as produced by Prokka, were clustered using Roary as outlined in Section 2.3.12. The output of the Roary clustering, using a 90% AA similarity cut-off, shows that within plasmids belonging to similar genotype (Table 12-1, Figure 12-2), there is significant discrepancy between the conservation of AA of certain CDSs (Figure 12-3). A very small portion of these differences in presence and absence can be explained by *de novo* annotation artefacts created by Prokka, as for example illustrated by the cleavage of the *areA* ORF compared to the pADAP reference, as described in Section 4.7. However, these artefacts can only account for a very small percentage of the diversity of CDSs within genotypes. The majority is caused by genetic diversity of these CDSs. It a clear example of genetic drift due to low selective pressure, as exemplified by the high diversity within the Sep encoding region (Chapter 5). This genetic drift can be interpreted as evolutionary distance between the plasmids, as mutation rate in plasmids are substantially higher than the chromosome (394). The plasmid still requires substantial time to accumulate enough mutations to account for AA identity to drop below the 90% cut-off used by Roary. Raw output obtained from Roary can be viewed in Supplementary Materials S.32, as well as a longer format plot showing labels for each ortholog cluster in Supplementary Materials S.38.

Using the Scoary toolset (267), as outlined in Section 2.3.13, the Roary output was analysed for statistically significant CDSs correlated to user defined classifiers such as ‘Chronic’, ‘Hyper’, ‘Mixed’ and ‘Non-path’ phenotype, based on bioassay data generated for each isolate (Section 12.4). Unfortunately, the Scoary toolkit was unable to determine any ORFs to be directly correlated to one of these phenotypical classifiers with enough statistical significance (empirical p-value ≤ 0.05). This is likely due to the highly diverse plasmids, sequenced for this study, coupled with the small sample size of some of these pADAP-type shown in Figure 12-2. Full outputs of the Scoary analysis are presented in Supplementary Materials S.39.

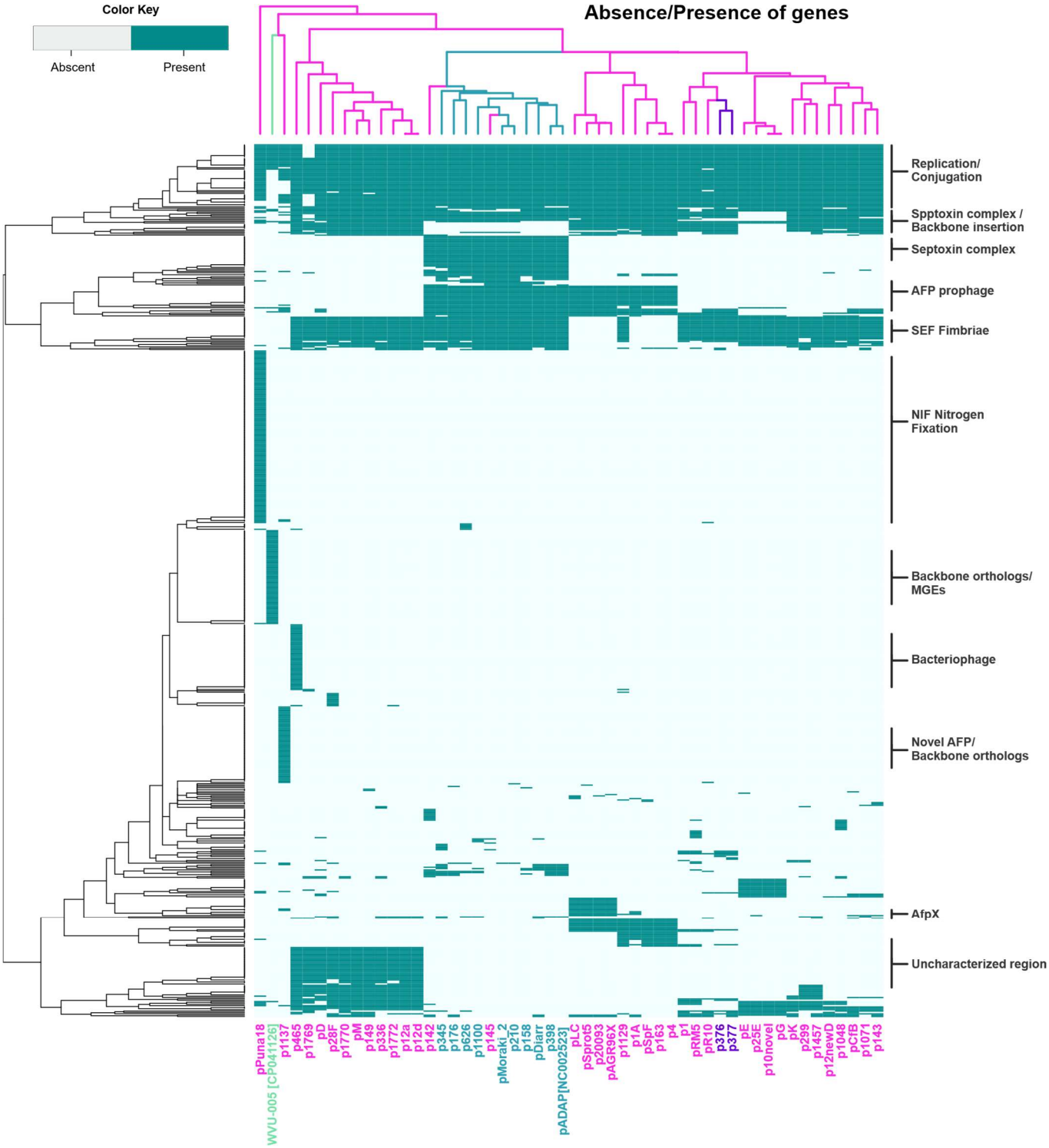


Figure 12-3: A visualization of the absence/presence matrix produced by Roary. On the X-axis are the 53 pADAP-type plasmids, and on the Y-axis the 528 orthologous gene groups. There are several unique regions observable in this plot. A larger plot with readable labels for each orthologous group on the Y-axis is provided in Supplementary Materials S.32. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens* and pink *S. proteamaculans*.

Based on the backbone conservation discussed in Chapter 4, and the variation of carried PAIs and GIs as described in the previous Section 12.1, as well as the disparity between CDSs in plasmids proposed to be of similar genotype as shown in Figure 12-3, it is clear that most plasmids have undergone substantial change through HGT facilitated indels and nucleotide mutation and replacement through recombination. This is also reflected in the nucleotide distance matrix presented in Figure 12-4. Looking at the meta-data obtained from the Roary clustering, as shown in Figure 12-5 and Figure 12-6, some key features of the pADAP-type plasmid family can be observed.

The plasmids belonging to the pADAP plasmid family contain at least 39 core genes, the majority of which make up the pADAP-type backbone. This cluster of core genes can also be observed in the block of Replication/Conjugation associated CDSs present in most plasmids show in the absence/presence matrix (Figure 12-3). On average there are approximately seven unique genes (i.e. encoding proteins of <90% AA similarity), per plasmid, although this number is slightly skewed by the inclusion of outliers such as pPuna18 (unique region 114680bp, containing 119 unique CDSs) and WVU-008-1 (unique region 28330bp, containing 30 unique CDSs).

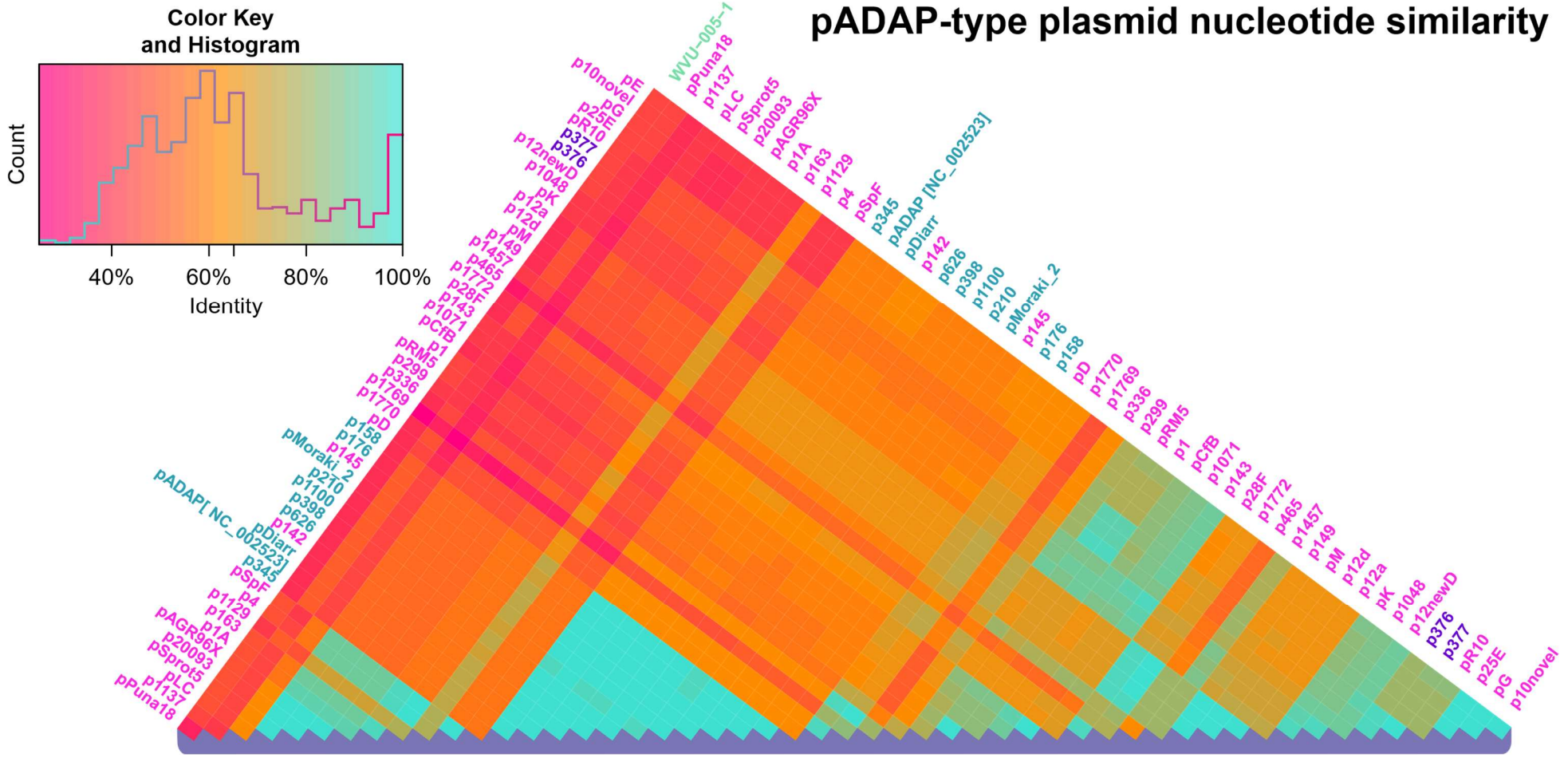


Figure 12-4: A DNA similarity matrix of the complete sequence of all 52 pADAP-like plasmids. Blue labelled strain names denote *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens* and pink *S. proteamaculans*.

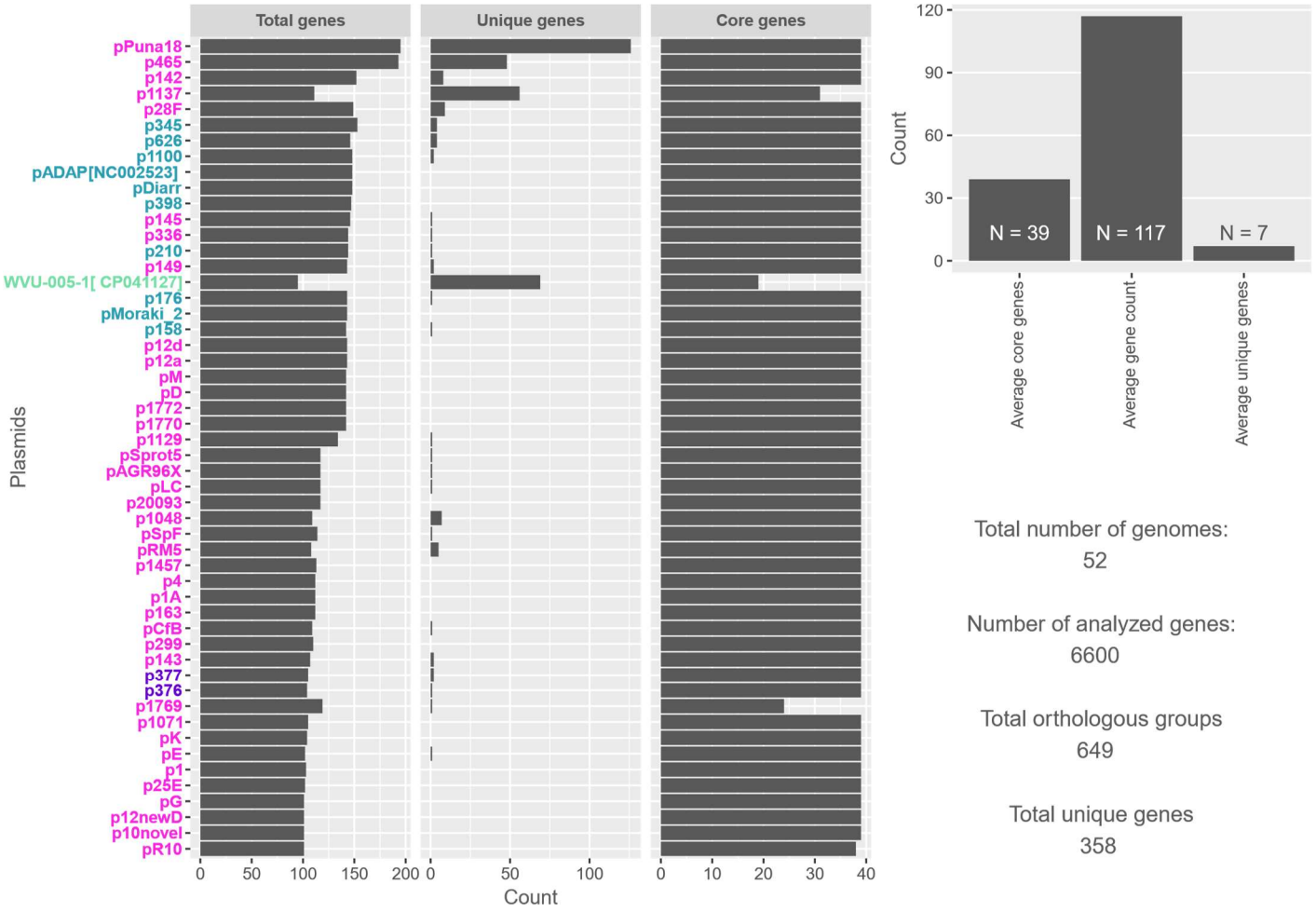


Figure 12-5: Graphs displaying general statistics of the plasmids assessed in this study. These statistics include total genes per plasmid, unique genes not found in rest of the set, genes in the set found in 95% of samples, average counts of genes, core genes and unique genes, and general statistics about the size of the entire CDS dataset. Blue denotes the *S. entomophila* strains, purple, *S. liquefaciens,* green, *S. marcescens* and pink, *S. proteamaculans*.

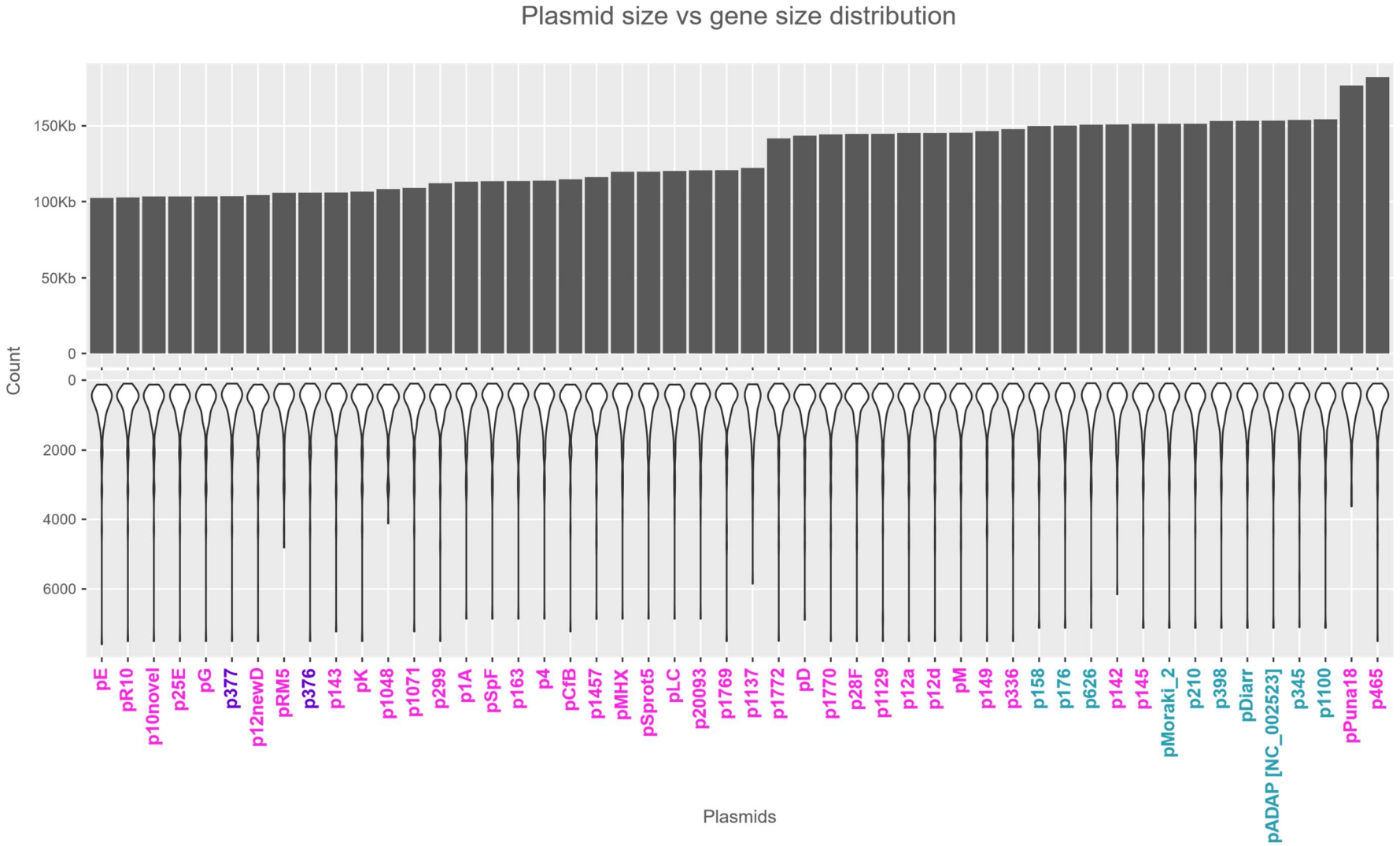


Figure 12-6: Plots depicting plasmid size (top), and violin plot representing CDS size distribution for each pADAP-type plasmid (bottom). Blue denotes the *S. entomophila* strains, purple, *S. liquefaciens* and pink, *S. proteamaculans*.

12.3 **Co-evolution of pADAP-type plasmids to chromosome of isolates**

As discussed in Chapter 4, the pADAP backbone could be referred to as SACA, as it has been found in numerous different species of *Serratia* and has been shown to contain several different types of GI and PAIs (Chapter 5 - Chapter 8). To determine whether these regions have been recently been acquired, or were a distant acquisition that underwent further speciation, the nucleotide bias, GC skew and codon usage of each plasmid was analysed. Methods used to perform these analyses are outlined in Section 2.3.14.

12.3.1 Nucleotide bias

One means of discerning if genetic elements are horizontally acquired or have co-evolved with their chromosomal hosts is by analysing the GC skew (121, 395). Most studies indicate that a lower GC% compared to the global average is indicative of HGT derived genetic material (396, 397). Amelioration of novel genetic material to its host mostly occurs through undergoing base mutations from weaker double hydrogen bound AT to stronger triple hydrogen bound GC nucleotides, in order to conform to its host GC% (398). Using this information, an attempt was made to show that the pADAP type plasmids most likely have been acquired over a long evolutionary distance and have co-evolved with their *Serratia* counterpart (as outlined in Section 2.3.14). Analysing the nucleotide bias of CDSs on the backbones of the plasmids, generated using the GCUA toolset (268), revealed that the GC% is approximately 49% for each *Serratia* species (Figure 12-7), whereas the chromosomal GC% for *S. entomophila* is approximately 51% and that of *S. proteamaculans* and *S. liquefaciens* is ~53%, as identified by Amy Vaughan [unpublished data]. As the WVU-005 dataset was added after this analysis was performed, the data for the pADAP-type plasmid WVU-005-1 was not included in the graph. The nucleotide bias for the CDSs found in the backbone of WVU-005-1 is A=25.1%, C=25.3%, G=26.2%, T=23.0% compared to a 58% chromosomal GC%.

The GC% of the assessed plasmids do not deviate much from their chromosomal counterparts. These results are not surprising when realizing that there are pADAP-type plasmid variants that are potentially plant associated (pPuna18), based on the carriage of plant related secondary metabolites and enzymes, or even mammalian associated (pWVU-005), therefore are unlikely to have crossed paths with isolates associated with insects, which make up the majority of the dataset assessed in this study. Instead, these results are indicative of speciation through co-evolution with the chromosomal components of the isolate and provide a clue towards determining whether the SACA has been deep-rooted *Serratia* associated element.

When assessing the nucleotide bias of CDSs of the entire plasmid (Figure 12-8), the *S. proteamaculans* based plasmids show significantly more variant nucleotide distributions. This shows that although the backbones of the pADAP family of plasmids have ameliorated to the chromosomal nucleotide bias over a potentially longer period of evolutionary time, some of the GIs and PAIs such as Sef, Sep and Afp, outlined in Chapter 5 - Chapter 8, have not undergone this process yet, and therefore have likely been recently acquired, although variations are slight. No change was observed in the *S. entomophila* pADAP plasmids, indicating that these plasmids are fully optimized to their chromosomal counterparts and haven't undergone substantial change in recent times. The WVU-005-1 plasmid was again not included in the figure, but displays little difference A=25.0%, C=25.7%, G=26.7%, T=22.6% relative to the nucleotide bias found in the plasmid backbone.

Raw nucleotide bias data, as well as the R-script used to generate their respective plots, are provided in Supplementary Materials S.40, S.40 and S.40.

12.3.2 GC skew of GIs and PAIs

When a GC skew is overlaid on top of the synteny map, it become clear which GI's and PAIs, or parts thereof, have a significant GC skew relative to the pADAP-type backbone (Figure 12-9). The RUF GI (discussed in Section 8.1), shown in red in Figure 12-9, shows a very low GC%, indicative of a more recently acquired sequence. Most notably, the area surrounding the orphan *sppA*/*sppB* (visualized in Figure 8-1, light red in Figure 12-9) reveals a substantial drop in GC% compared to the rest of the plasmid, which is to be expected as they are remnant paralogs, meaning there is very low selective pressure on them, if any at all, allowing for higher mutation rates without affecting pathogenicity or plasmid stability.

Areas surrounding Afp17/18 also show lower GC%, indicative of higher mutation rates. These findings are in line with a recombination hotspot. Higher rates of mutations are also a known feature to have for organisms in a predator-prey relationship (399), as is the case with pADAP bearing *Serratia* isolates and their grass grub larvae prey. Routine mutation are key in speciation, and genomic drift (394).

When comparing plasmids derived from *S. entomophila* such as Type A (pADAP), F (p1137) or G (pPuna18) to a *S. proteamaculans* type such as Type B (pAGR96X) in Figure 12-9, the divergent backbone region outline in Section 4.9, is also visually represented by a drop in GC% at the ~27 Kb point only found in the *S. proteamaculans* plasmids such as Type B in Figure 12-9. This drop overlays with the inserted region and shows that this region is a more recently acquisition in *S. proteamaculans* pADAP-type plasmids, as opposed to a recent loss by *S. entomophila* and *S. marcescens* pADAP-type plasmids.

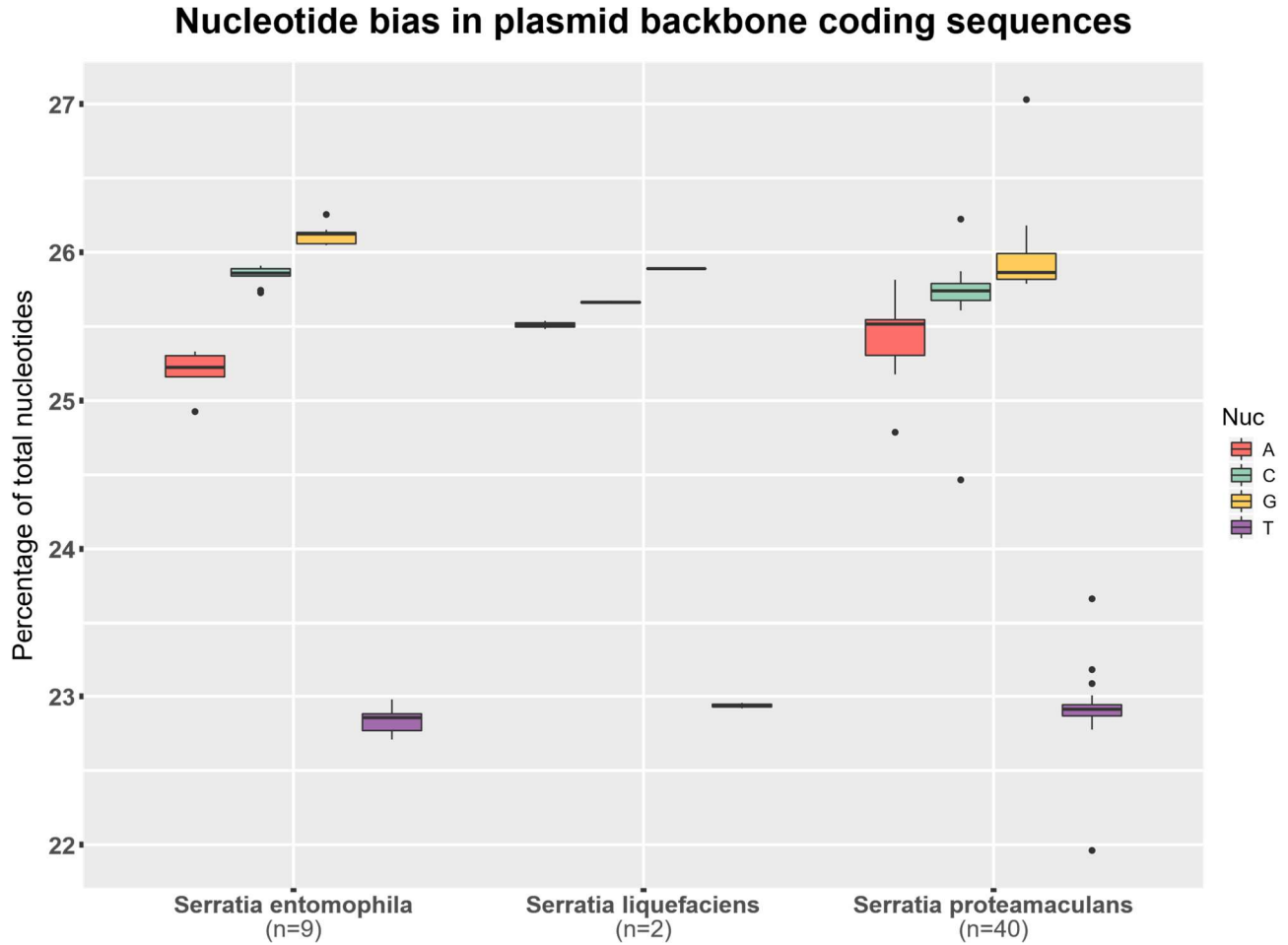


Figure 12-7: Whisker plot showing the nucleotide biased in the pADAP-type plasmid backbone CDSs. Although it seems as if *the S. proteamaculans* have more divergence nucleotide bias, this can also be explained by the larger number of *S. proteamaculans* pADAP-type plasmids in this dataset. The *S. proteamaculans* group also contains two outlier strains, 1769, which is missing portion of the conjugation region and the p465 plasmid backbone, which has a Mu-like phage inserted into it.

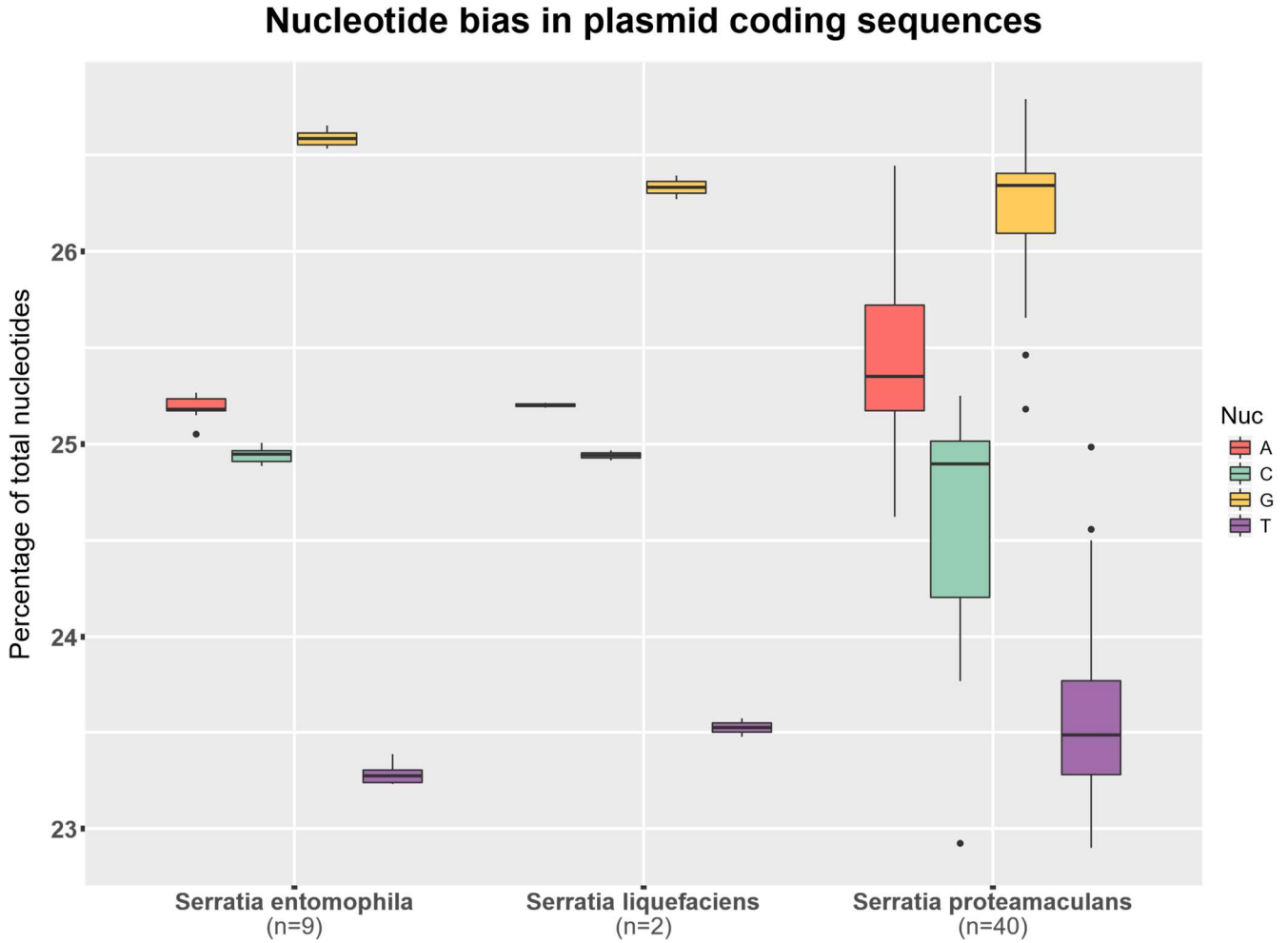


Figure 12-8: Whisker plot showing the nucleotide biased over all pADAP-type plasmid CDSs. Although it seems as if the *S. proteamaculans* have more divergence nucleotide bias, this may reflect the larger number of *S. proteamaculans* pADAP-type plasmids in this dataset.

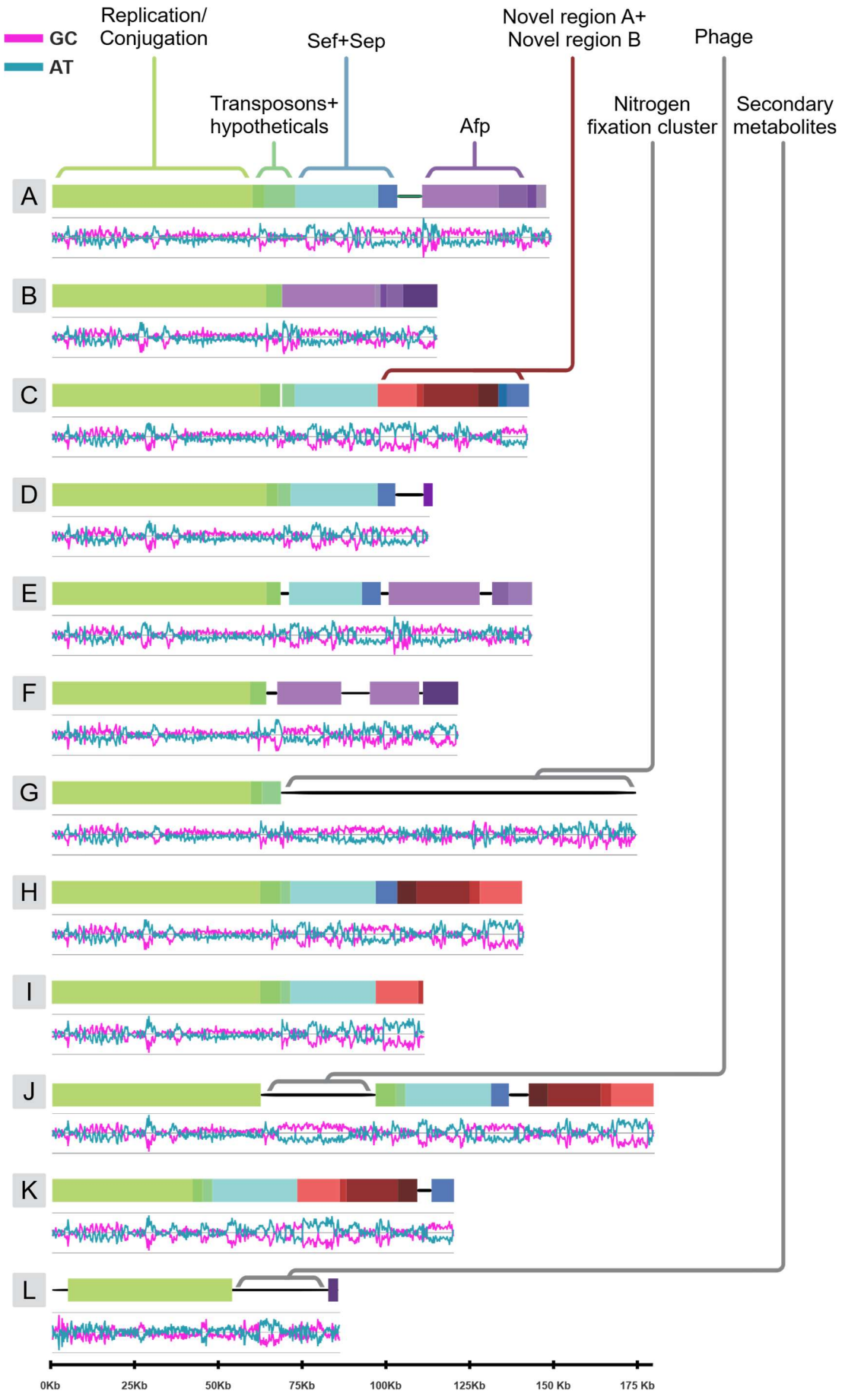


Figure 12-9: A GC distribution plot, overlaid on top of a synteny map for a representative of each of the 12 pADAP genotypes.

12.3.3 Codon usage analysis

Another way of assessing whether horizontally acquired MGEs, such as these pADAP plasmids, frequently change hosts or whether they have co-evolved with their host over a long evolutionary time, is by analysing the amino acid frequency and codon usage bias (400). As bacteria reside in different micro milieus, their nutritional intake will be substantially different. Meaning, for some bacteria, certain amino acids can be more or less abundant than others. Using the GCUA toolset (268) as outlined in Section 2.3.14, all translated CDSs obtained from the pADAP-type plasmids were analysed for amino acid frequency, codon usage, and stop codon bias. Figure 12-10 displays that amino acids residues such as Leucine (Leu) and Alanine (ALA) are heavily favoured over Tryptophan (Trp) and Cysteine (Cys) for example. As shown, there is little variation between the different plasmids despite being associated with different target organisms such as the pADAP carrying *S. entomophila* and pAGR96X bearing *S. proteamaculans* being insecticidal plasmids associated with grass grub larvae, the *S. marcescens* associated WVU-005-1 being associated with mammals, and the *S. proteamaculans* isolate carrying pPuna18, which encodes for a nitrogen fixation cluster, most likely being a plant associated plasmid.

The amino acid frequencies themselves are not noticeably different from frequencies observed in other organisms, as there is an innate bias in amino acid usage such as an abundance of Leucine and Serine residues in almost all types of organisms (401, 402). Leucine residues have a high propensity to be buried due to their non-polar hydrophobic qualities: their main function is maintaining a stable hydrophobic core and are rarely directly involved with protein function. Serine is a lightly charged residue found on the inside and surface of many proteins, most notably in active sites of proteases, lipases etc. The frequency of amino acids can therefore not be affected too much in order to be able to make stable protein products. The less abundant residues such as the mentioned Tryptophan and Cysteine both serve specialist functions and are used less frequently across all organisms. Tryptophan is hydrophobic residue with non-polar side chain, it helps in forming sharp turns in tertiary structures and Trp plays a crucial role in anchoring membrane proteins. Cysteine is generally considered a polar and hydrophilic residue. The Cys residue contains a sulphur group which is able to form strong disulphide bonds in proteins, increasing rigidity and provide resistance from proteolytic activities from enzymes.

The codon usage used to make mRNA however has been shown to change from species to species within the same genus (400). Plasmids containing CDSs that heavily bias towards certain types of transfer RNA (tRNA) could deplete availability of needed tRNAs and reduce chromosomal mRNA production, potentially resulting in lower fitness of the cell (129, 154). Therefore, codon usage could result in incompatibility between plasmid and chromosome. A heatmap configuration of the codon biased expressed in relative synonymous codon usage frequency (RSCU) (403) is presented in Figure 12-11 for the pADAP-type backbone, and Figure 12-12 for the entire pADAP-type plasmids. The RSCU score is the observed frequency of amino acids compared to the expected frequency of all synonymous codons being present equally.

Through the assessments of the heatmaps depicted in Figure 12-11 and Figure 12-12, it was observed that there is a high affinity towards CGC and low affinity towards AGA and AGG to encode for an Arginine (Arg) residue, as well as a high affinity towards CUG and low affinity towards CUA to encode for a Leucine (Leu) residue. Other than these two examples, most codons are relatively equally used across all plasmids assessed in this study. No significant differences are observed between codons used in the pADAP-type backbones compared to the CDSs of the entire plasmids.

Assessments of the stop codon bias in pADAP-type plasmids, revealed that all plasmids exhibit the UAA, UAG, UGA stop codons of which UAG is used the least prevalent. Models by Povolotskaya et al. (404) show that UAA codons are overly present in bacteria with low GC% and UGA in GC rich bacteria. They also showed that UAG is negatively selected on based on substitution rates. With the pADAP-type plasmids showing an average GC content of 49% (Figure 12-7), as well as having a chromosomal GC% of 51%, 53% and 53% for *S. entomophila*, *S. proteamaculans* and *S. liquefaciens* respectfully as assessed by Amy Vaughan [unpublished data] (Section 12.3.1), they conform to the predicted stop codon distribution observed in the Povolotskaya model.

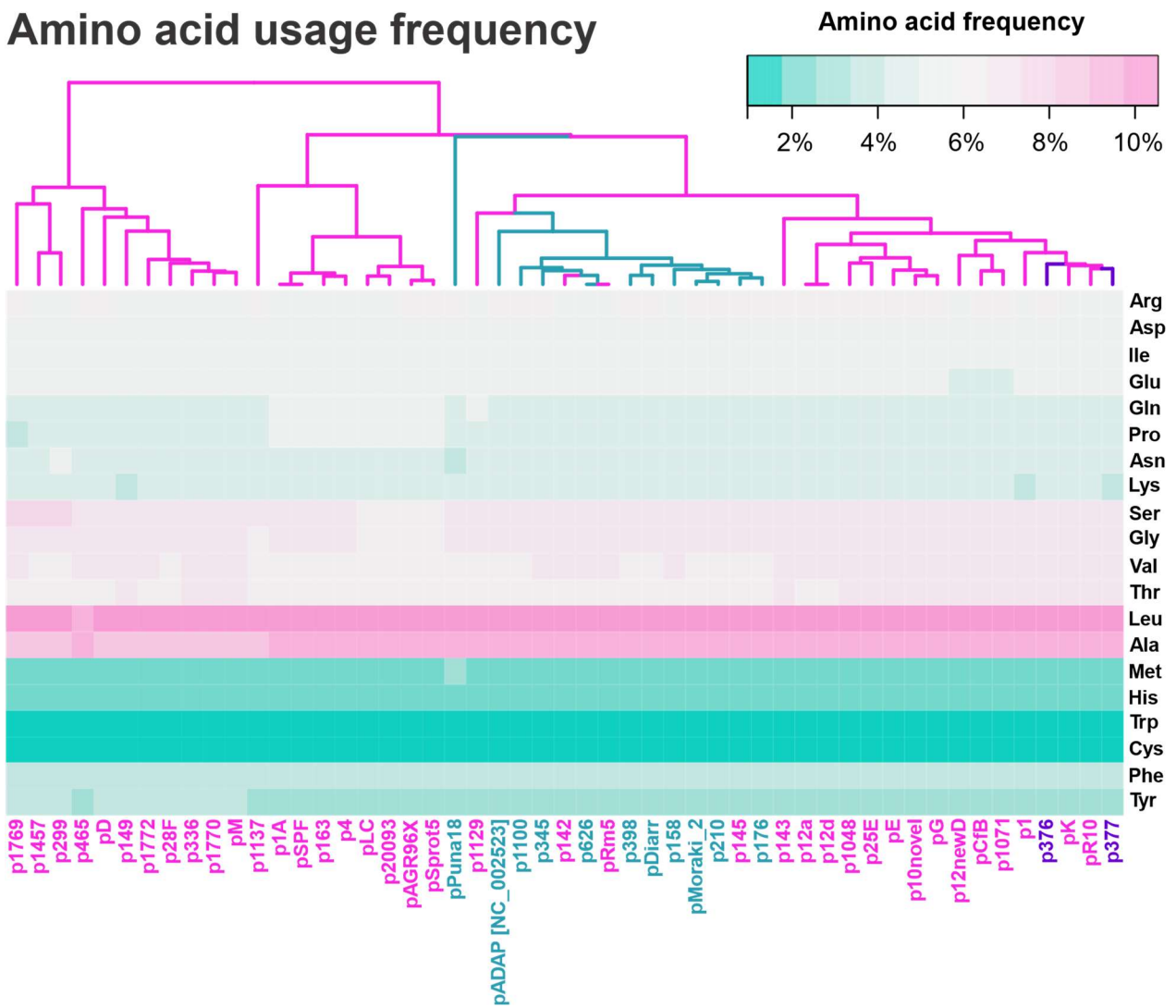


Figure 12-10: Average frequency of amino acids used in each pADAP-type plasmid. Blue labelled strain names are *S. entomophila*, purple *S. liquefaciens*, pink *S. proteamaculans*.

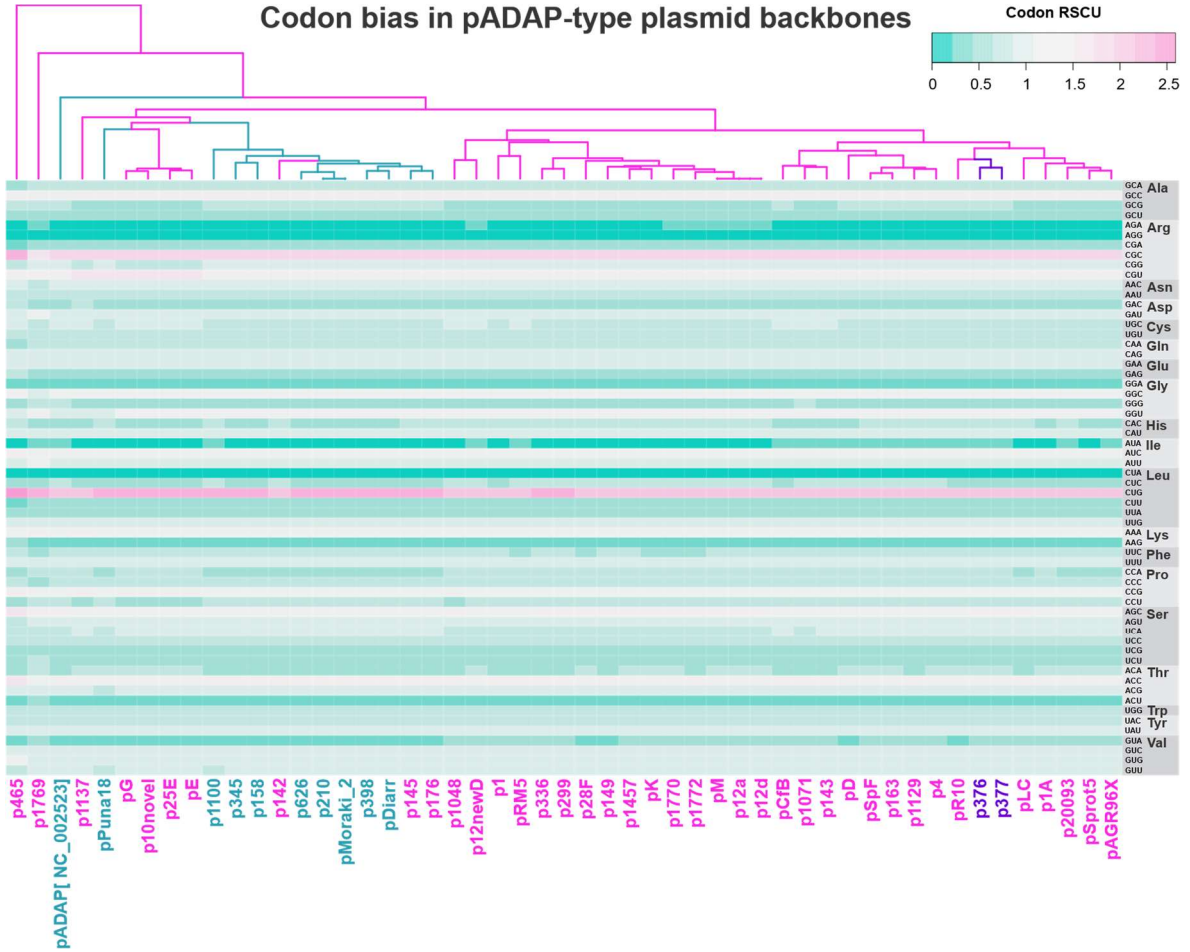


Figure 12-11: Codon usage plot for pADAP-type plasmid backbone CDSs. This plot reveals the codon biased for the plasmid backbones. Apart from the two outliers 1769, which is missing the conjugation region and the 465 plasmid backbone, which has a Mu-like phage inserted into it, there are no noteworthy differences in codon bias. Blue labelled strain names are *S. entomophila*, purple *S. liquefaciens*, pink *S. proteamaculans*.

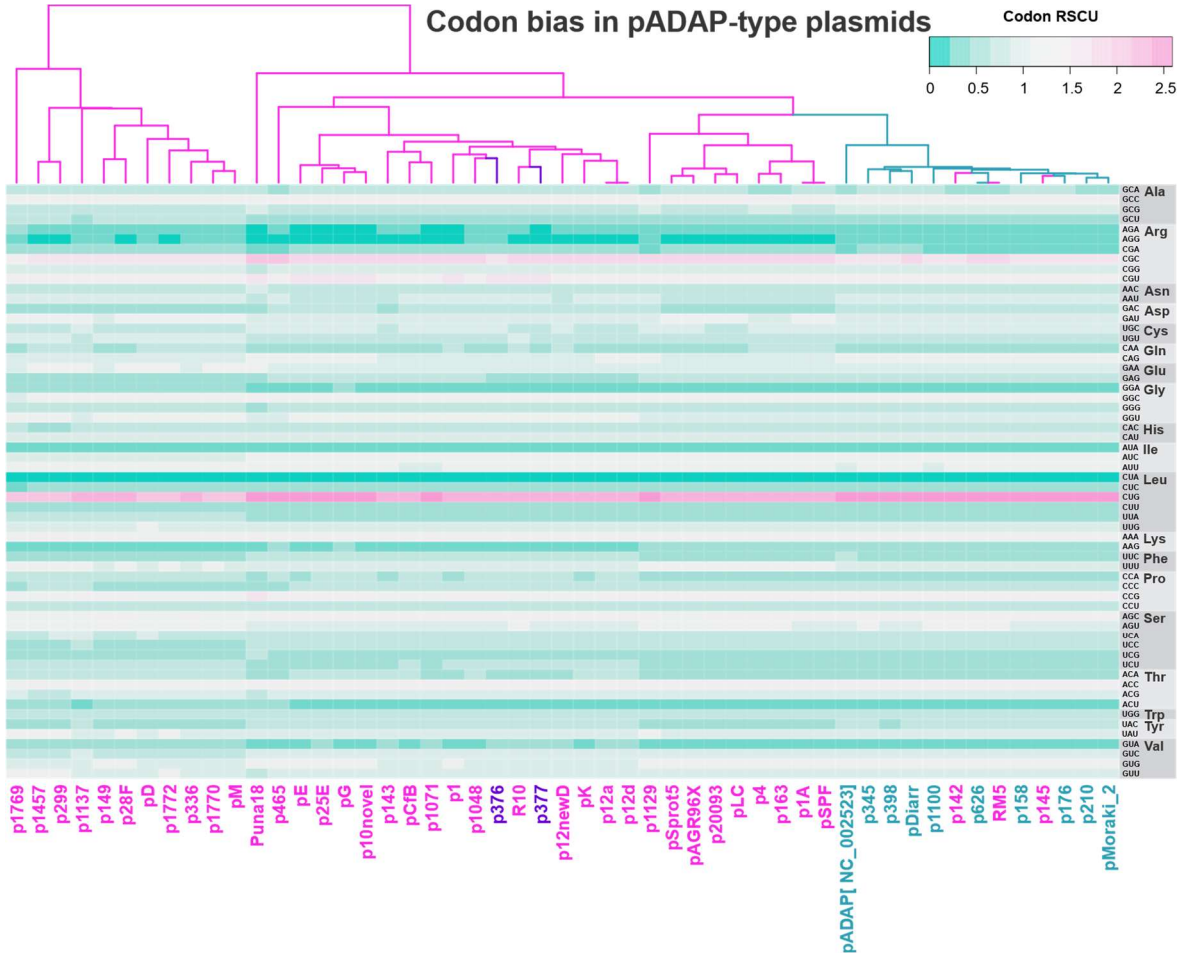


Figure 12-12: Codon usage plot for pADAP-type plasmid backbone CDSs. This plot reveals the codon bias for the pADAP-type plasmids. Blue labelled strain names are *S. entomophila*, purple *S. liquefaciens*, pink *S. proteamaculans*.

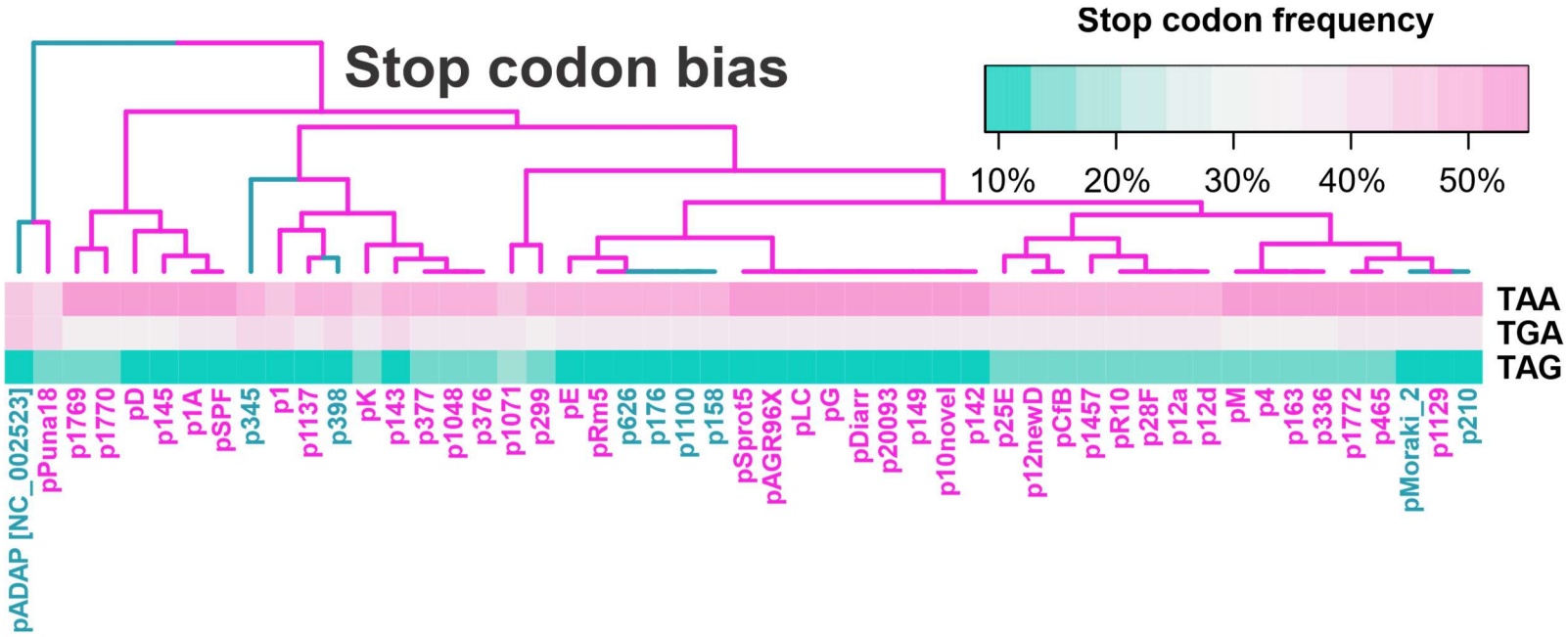


Figure 12-13: Stop codon bias analysis performed on all pADAP-type plasmids. Blue labelled strain names are *S. entomophila*, purple *S. liquefaciens*, pink *S. proteamaculans*.

12.4 **pADAP-type plasmid and chromosomal host correlation**

As exemplified in the previous Sections 12.1, Section 12.2 and Section 12.3, a hypothesis was formed that the pADAP family of plasmids were once a SACA, comprising the elements found on the pADAP backbone, which has accrued different PAIs and GIs over a long evolutionary timeframe. Through this time, older elements such as the Sep GI have potentially undergone large amounts of mutations leading to further speciation and resulting in different bioactivity in grass grub larvae, presented in Table 3-1. This mutation through low selective pressure might have been a result of certain subspecies moving out of the grass grub associated milieu and entering other ecosystems.

12.4.1 Phylogeny

As outlined in Section 4.1Chapter 4, RepA is the core replication element needed for replication of pADAP-type plasmids. There are RepA orthologs found on numerous non-pADAP plasmids as depicted in Figure 4-1. Therefore, the RepA can be used to display the evolutionary distance between pADAP-type plasmids. The nucleotide alignment-based ML tree, depicted in Figure 12-14, displays clear clustering of the S. entomophila pADAP plasmids as a separate group.

In Section 12.3, an attempt was made to show that despite the potential handcuffing described in Section 4.2, the main limitation behind pADAP-plasmids recombining and optimizing PAIs by exchanging them out for more efficient PAIs such as AfpA or Afp for AfpX, is most likely due to the limited interaction these isolates have with one another. To show this, a 16s based phylogenetic tree was correlated to a phylogenetic tree generated using 13 core pADAP-type backbone genes (as outlined in Section 2.3.15) that were found in all pADAP-type plasmids analysed in this study. Using a tanglegram approach, the two trees were linked to one another. This analysis was undertaken to see if plasmids would cluster to members of the same genotype as shown in Section 12.2, based on only these 13 core backbone ORFs, as well as see if these clusters show similarities to the clustering of the chromosome.

The 16s ML tree tangled to the pADAP-type core backbone gene super tree is presented in Figure 12-15. The figure reveals that all chronic disease inducing pADAP plasmids originating from *S. entomophila* (as well as the two pADAP plasmids found in the *S. proteamaculans* isolates 142 and 145) cluster together in the plasmid backbone tree, and correlates nicely with the clustering of the *S. entomophila* 16s tree. This shows that the pADAP plasmid is a plasmid that is most notably associated with *S. entomophila* rather than *S. proteamaculans*, with the *S. proteamaculans* 142 and 145 isolates being the exception.

In the instance of p142 and p145, these are indicative of the potential for these pADAP plasmids to be transferred to different species within the *Serratia* genus as is shown experimentally in Section 12.4.2.

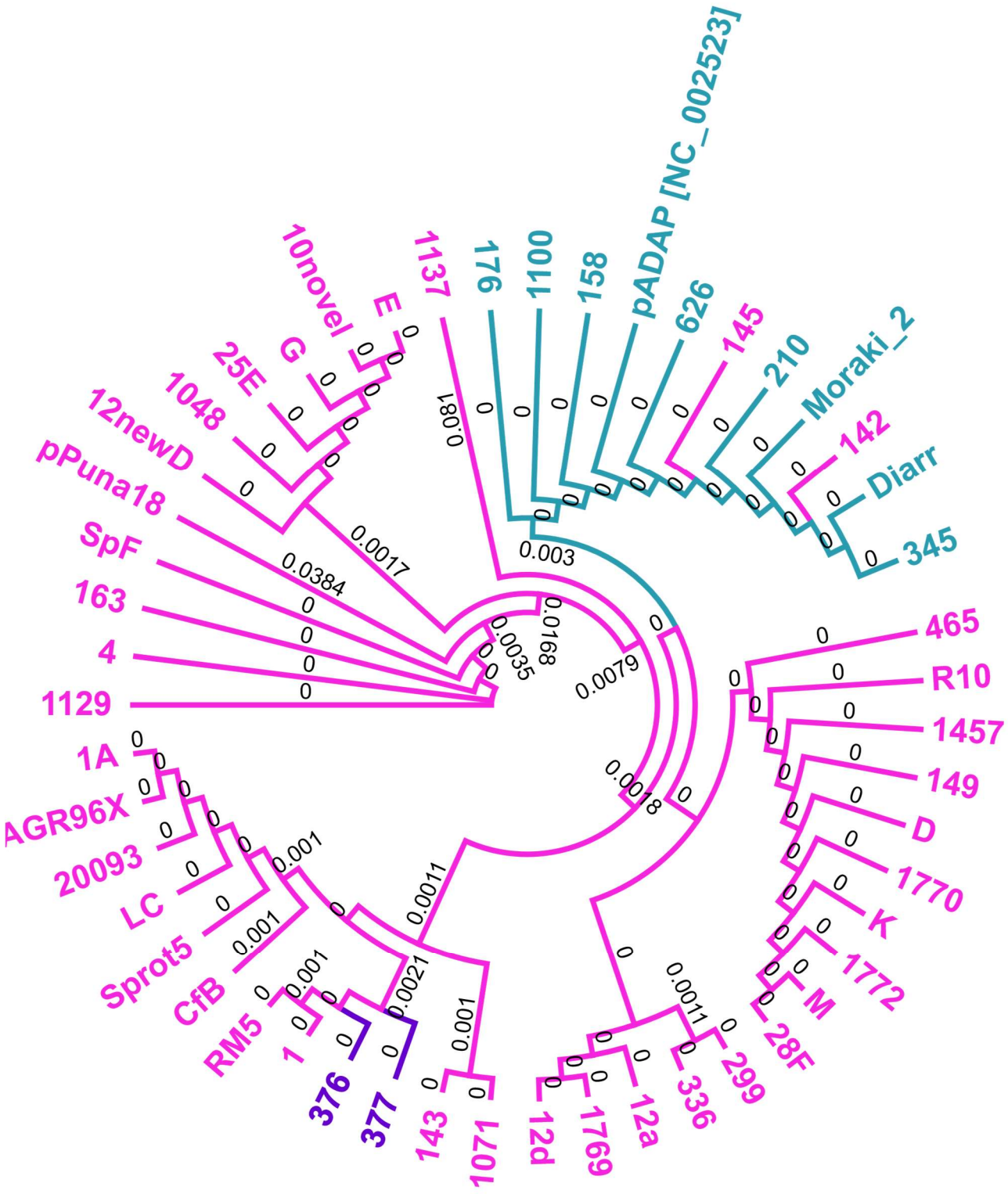


Figure 12-14: A *repA* ML tree. Blue labelled strain names are *S. entomophila*, green S. marcescens, purple *S. liquefaciens* and pink *S. proteamaculans*.

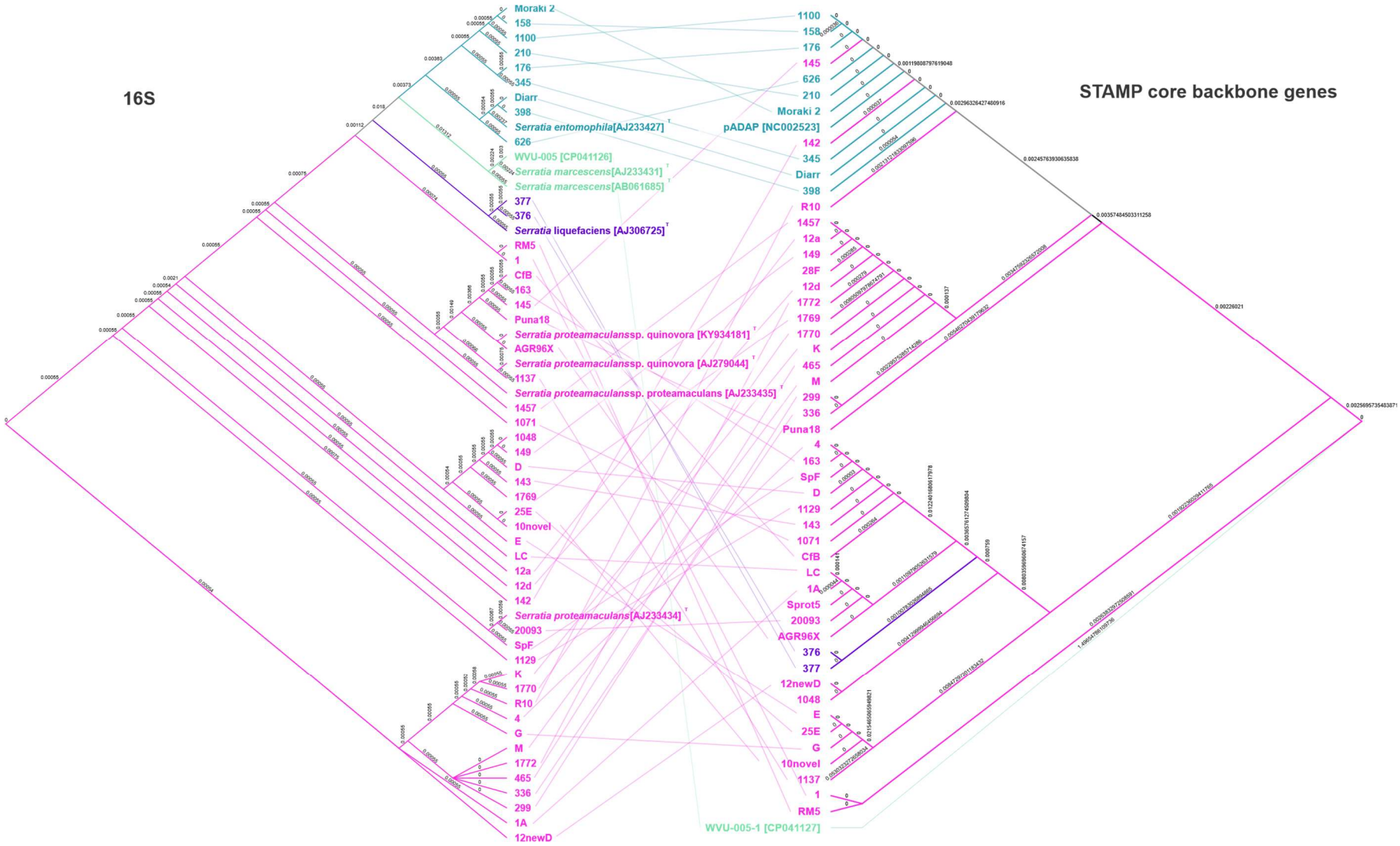


Figure 12-15: 16s ML tree tangled to the pADAP-type core backbone gene supertree. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens* and pink *S. proteamaculans*.

No *S. proteamaculans* associated pADAP-type plasmid was observed in a *S. entomophila* isolate. The two *S. liquefaciens* isolates 376 and 377 cluster into their own 16S group, however their pADAP-type plasmids appear to be *S. proteamaculans* derivatives, potentially indicating that the *S. liquefaciens* isolates have acquired the plasmids later than *S. proteamaculans*, and potentially through an *S. proteamaculans* donor.

The outlier *S. marcescens* strain WVU-005, as expected, clusters with the *S. marcescens* 16s reference sequences. However, the WVU-005-1 plasmid is the most distinct plasmid based on the 13 backbone ORFs, and may indicated that the pADAP-type backbone potentially has been around of a very long time, and thus are only seeing a distinct branch of pADAP-type plasmids in this study.

12.4.2 Conjugation of pADAP-type plasmids

Based on the results presented in the previous section in this chapter, the reason behind the divergence of pADAP-type plasmids, and the lack of convergence towards one optimal insecticidal plasmid is most likely the lack of opportunities to exchange plasmids, as well as limitations in plasmid recombination. The outlier plasmids WVU-005-1 and pPuna18, found in *S. marcescens* and *S. proteamaculans* respectively, are great examples of the SACA having acquired GIs with no apparent insecticidal functionality, and therefore illustrate speciation and exemplify that the pADAP-type backbone has co-evolved with the *Serratia* genus over a long evolutionary timeframe.

Grkovic et al. (85), showed that conjugation from the *S. entomophila* into other *Serratia* species was possible, and that the conjugated plasmid was able to induce the pathogenic response in grass grub larvae in a bioassay. It also showed that the *S. liquefaciens* 377 isolate (wrongly labelled *S. proteamaculans* in the article), upon conjugation, lost its original plasmid, which this study showed to be a pADAP-type plasmid. This loss is likely due to the self-incompatibility of pADAP-type plasmids (Section 4.2). In another article by Glare et al. (86), it was shown that the pADAP plasmid could be conjugated into other members of the Enterobacteriaceae family, albeit with an additional mobilization cluster cloned into the plasmid. These examples show that pADAP has very little problem transferring from *S. entomophila* to *S. proteamaculans*.

To confirm whether it's also possible to transfer from *S. proteamaculans* to *S. entomophila*, several triple tagged pADAP-type plasmids were conjugated to the pADAP heat-cured mutant of the *S. entomophila* A1MO2 typestrain, labeled 5.6 (81), as well as a plasmid-free *S. proteamaculans* strain, labeled 3041 or Tukino (81).

The *S. proteamaculans* pADAP plasmid p145 as well as the *S. entomophila* pADAP plasmids p210, p398 and p1100, were triple tagged by recombining Cm^R, Sp^R and Km^R cassettes into the plasmid backbone, Sep region and Afp region respectively, as further discussed in Section 2.2.14. The *S. proteamaculans* pADAP plasmid p142 as well as the pADAP-type plasmids p149 and p1129 were only tagged in the backbone with a Cm^R cassette. The *S. proteamaculans* pU143 plasmid does not encode for Afp therefore was only tagged in the backbone and Sep region with Cm^R and Sp^R respectively.

Table 12-2: Results of trans conjugating tagged plasmids to plasmid free strains 5.6 and 3041.

Species		Isolate	Plasmid	5.6*	3041*
Serratia	*entomophila*	210	p210All	✓	✓
Serratia	*entomophila*	398	p398All	✓	✓
Serratia	*entomophila*	1100	p1100All	✓	
Serratia	*proteamaculans*	142	p142BB[1]	✓	✓
Serratia	*proteamaculans*	143	pU143BS[2]	✓	✓
Serratia	*proteamaculans*	145	p145All	✓	✓
Serratia	*proteamaculans*	149	p149BB[1]	✓	✓
Serratia	*proteamaculans*	1129	p1129BB[1]	✓	✓

* isolates were transformed with the Tc^R pACYC184ΔCM to facilitate screening

BB = Backbone tagged, BS = Backbone + Sep tagged, ALL = Backbone + Sep + Afp tagged

[1] Plasmids were only Cm^R tagged in the backbone

[2] Plasmid does not contain Afp, therefore was only Cm^R and Sp^R tagged in backbone and Sep region respectively

All results were validated using Box-PCR profiles (Section 2.2.6), and Kado and Liu (Section 2.2.10) to determine the donor plasmid was conjugated into the proper recipient cell [unpublished data]. Occasionally the pACYC184ΔCM vector transferred to the donor cell, but through the double validation using Box-PCR and mega plasmid visualization, it was possible to define the correct trans-conjugants and validate conjugation from one species to another. Bioassays performed with several transconjugants also showed that pathogenicity was transferred, as previously shown by Grkovic et al. (85) and Glare et al. (86) (Table 12-3). Of note, higher virulence was noted in the *S. proteamaculans* 145 WT as opposed to *S. entomophila* 5.6 or *S. proteamaculans* 3041 carrying the p145 plasmid (Table 12-3). Similar observation was made for the *S. proteamaculans* 210 WT compared to the *S. entomophila* 5.6 carrying the p210 plasmid, however the *S. proteamaculans* 3041 carrying p210 mutant produced similar bioassay results. These findings suggest the presence of a chromosomal component in the bioactivity towards grass grub larvae.

Table 12-3: Bioassay results of transconjugants of *S. entomophila* (210 and 1100) and *S. proteamaculans* (145) based pADAP plasmids conjugated into the plasmid free *S. entomophila* (5.6) and *S. proteamaculans* (3041 / Tukino) type strains.

Strain	Disease%			Mortality%			Affected%		
Control	2.4	±	1.7	1.2	±	1.2	3.6	±	2.0
5.6									
3041									
145	95.8	±	4.2	4.2	±	4.2	100.0	±	0.0
5.6 + p145_ALL	66.7	±	14.2	0.0	±	0.0	66.7	±	14.2
3041 + p145_ALL	66.7	±	14.2	8.3	±	8.3	75.0	±	13.1
210	90.9	±	6.3	4.5	±	4.5	95.5	±	4.5
5.6 + p210_ALL	75.0	±	13.1	0.0	±	0.0	75.0	±	13.1
3041 + p210_ALL	91.7	±	8.3	0.0	±	0.0	91.7	±	8.3
1100	87.5	±	6.9	8.3	±	5.8	95.8	±	4.2
5.6 + p1100_ALL	75.0	±	13.1	8.3	±	8.3	83.3	±	11.2

These results, combined with the results shown by Grkovic et al. (85) and Glare et al. (86), show that there is no limitation behind pADAP plasmid transfer within the *Serratia* genus apart from the physical location of donor and recipient cells carrying different pADAP-type plasmids.

Chapter 13 Summary of the pADAP family of plasmids and their features

When combining the results from all previous chapters, including the results presented in this chapter, the data revealed an abundance of evolutionary traits such as mutation, HGT, acquisition and deletion of genetic material, DNA shuffling and inversion of GIs occurring in these pADAP-type plasmids. This reveals that natural selection is allowing plasmids to evolve in different directions and potentially enabling their bacterial hosts to thrive in different niche environments. Although the initial goal of this study was to define correlation between presence and absence of major pathogenicity associated islands such as Sep and Afp PAIs and the bioactivity the isolates have towards *C. giveni* larvae, the results of this work revealed evidence towards genomic drift of plasmids, and indirectly chromosomes, to optimize bacterial fitness in numerous ways that may reflect an arms races for bacterial dominance.

As a summary of the results accumulated through this study, a figure was generated (Figure 13-1). In this figure the genetic types of each identified PAI and GI are shown using a colour key shown in the top left of the figure. This figure encompasses all the information presented in previous chapters. With reference to Figure 13-1, it is evident that there is a high degree of plasmid diversity.

A proposed model of the evolution for the pADAP plasmid family is depicted in Figure 13-2. This figure is meant to illustrate the hypothesized events that led to the formation of all the pADAP genotypes outlined in Section 12.1. As previously touched on, it's most likely a Serratia associated capturing apparatus, that consists of portions of the replication and the conjugation regions of the pADAP backbone, has co-evolved with different Serratia species and isolates for a long duration. It's unclear how long, as plasmids do not conform to fixed rates of mutation and often allow for HGT of large GIs or small MGEs and have lower selective pressure. In reference to Figure 13-2, this model assumes that regions observed more frequently, have been acquired prior to the ones that appear in smaller sub-populations.

There is very little correlation between presence or absence of certain GIs and the phenotype produced by the bacteria, nor is there any correlation between geographical location and type of pADAP-type plasmid carried by the isolate. The only conserved group of plasmids are the pADAP plasmids found in *S. entomophila*.

The pADAP carrying *S. proteamaculans* isolates 142 and 145 however, although carrying a highly conserved pADAP plasmid, do not have the same bioactivity as pADAP bearing *S. entomophila* isolates, indicative of a chromosomal aspect to pathogenicity as well. Further study into the chromosomal aspect is currently being undertaken by Amy Vaughan and will hopefully shine more light on the intricate mechanics surrounding proliferation, colonization and pathogenicity of pADAP-type plasmid bearing bacteria.

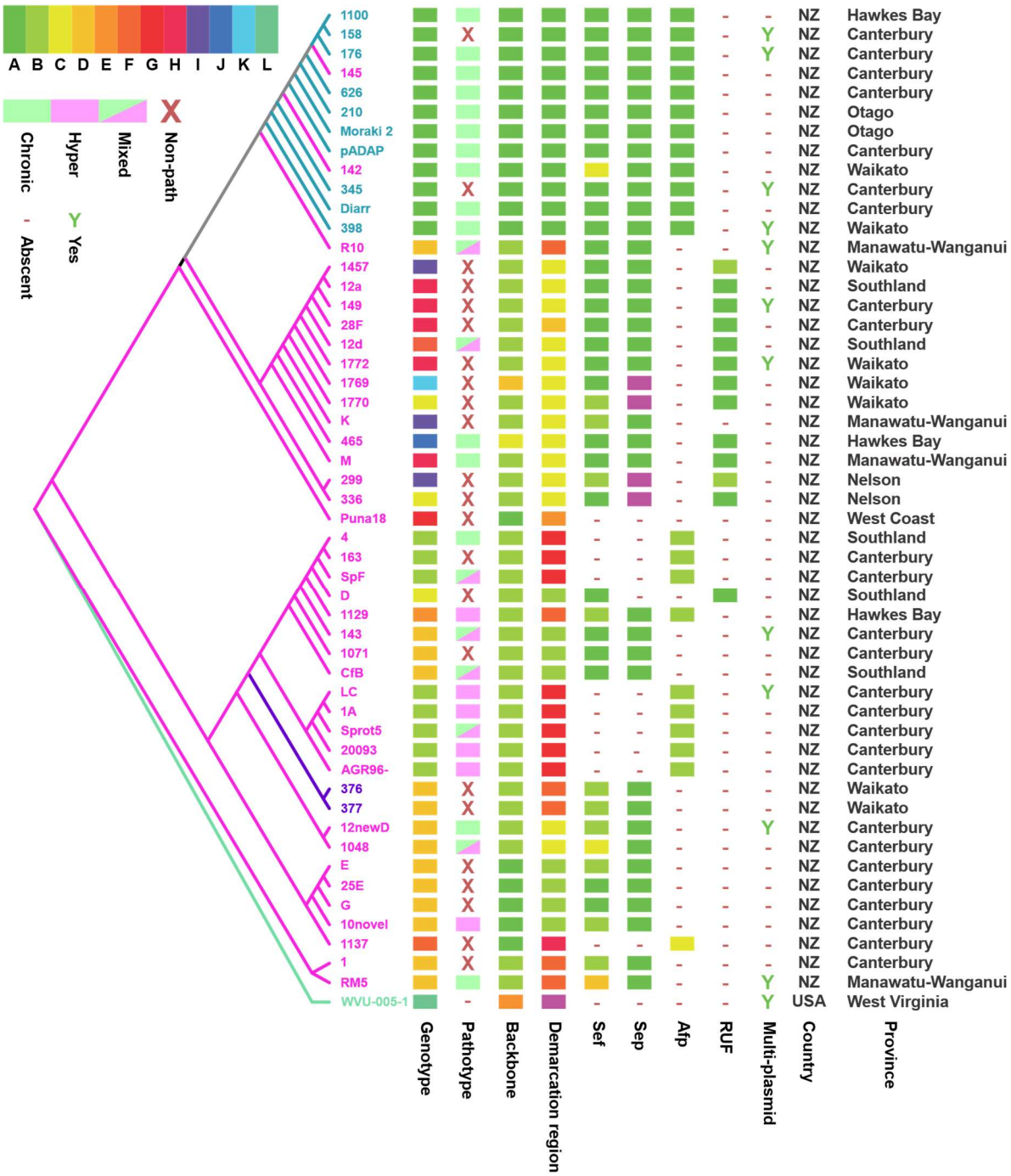


Figure 13-1: A high level overview of all the data accrued throughout the course of this study. The backbone tree shows the genetic similarity of plasmids based on 13 core backbone genes (Section 12.4). Each column represents the specific type of genetic makeup as described in previous chapters, Genotype (Section 12.1), Pathotype (Section 3.2), Backbone (Section 4.10), Demarcation region (Section 4.12), Sef (Chapter 6), Sep (Chapter 5), Afp (Chapter 7), RUF (Section 8.1), Country (Section 3.3). Due to large range of divergent Sep, only Sep bearing (green) and degenerate (purple) are shown. Additional data such as presence of additional plasmids in bacterial cell, and province from which isolate was obtained are also presented. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens* and pink *S. proteamaculans*.

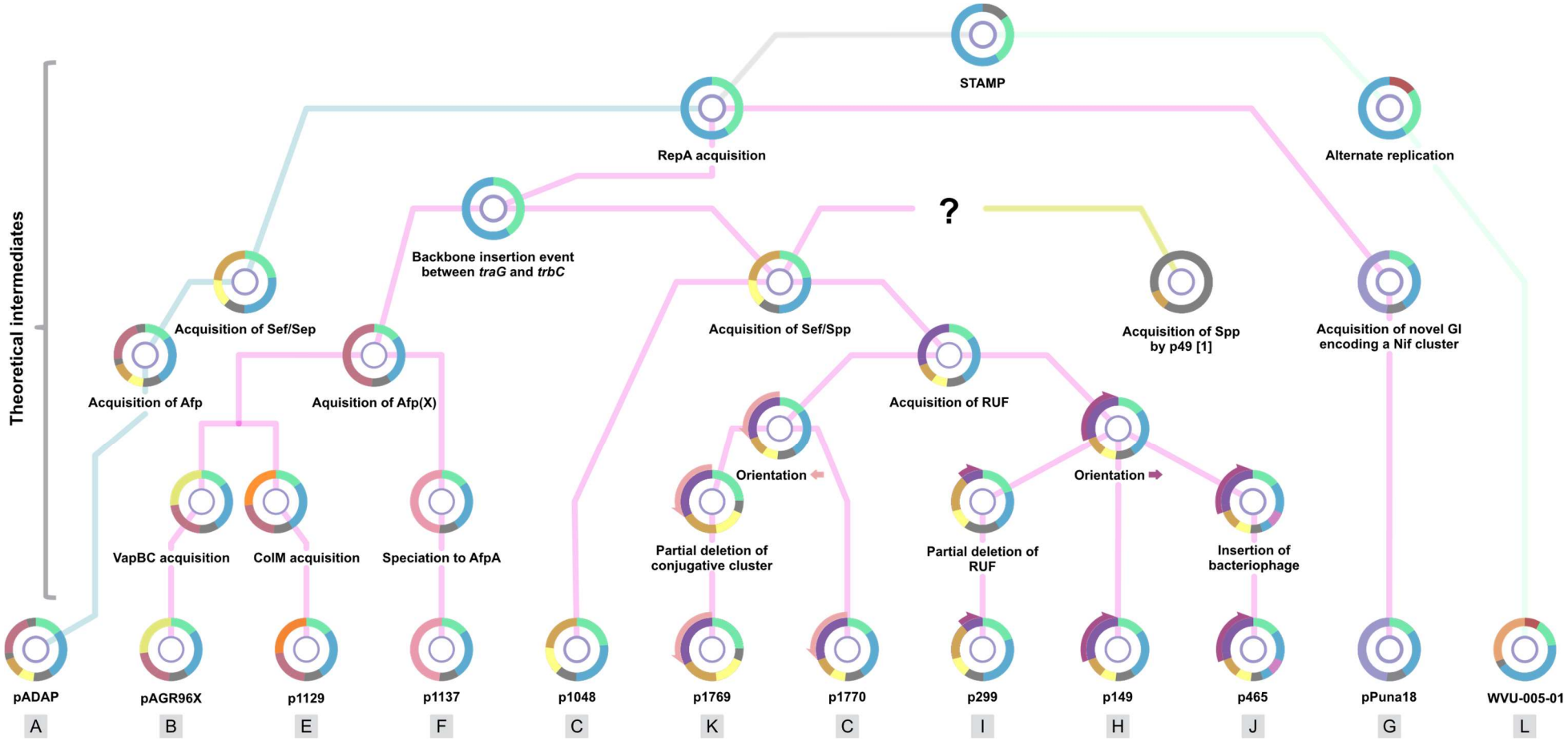


Figure 13-2: A model of the evolutionary speciation of the pADAP family of plasmids. This model is based on findings of this study and the assumption that regions shared more frequently are result of one shared event, and regions shared less frequent have happened subsequently. Blue branches are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens,* pink *S. proteamaculans*, orange *Y. frederiksenii*.

13.1 **The *Serratia* supergenome**

Plasmids allow diverse taxa to obtain novel genetic material. Plasmids shared between different taxa, also referred to as communal plasmids, are part of what is called the supergenome (117). The supergenome consists of shared genetic sets of MGE that can be exchanged between a communal pool of bacteria, and sometimes between several pools in different hosts or environments.

The pADAP plasmid family analysed in this study has at one point been part of this supergenome but now act more as secondary chromosomes or chromids. Instead several isolates analysed in this study carry additional plasmids, or do not carry a pADAP plasmid at all and instead carry other plasmids that might fulfil this role of actively partaking in the supergenome. Some of these non-pADAP plasmids carry pADAP elements such as the Sep encoding region found on a plasmid of *Y. frederiksenii* 49, identified by Dodd et al. (90) and discussed in Section 5.1.

Other non-pADAP plasmids such as the *S. entomophila* pD1 and the *S. proteamaculans* p591, pMan4 and pSm1a were also found to carry Sep as well as Sef elements. The p591 also carried a seemingly functional bacteriophage encoding region, also found to be encoded on the *S. proteamaculans* based pADAP-type plasmid p465 (Section 8.5). Other elements such as the shared VapBC element with non-pADAP plasmid p220-1 (Section 11.3) were identified as well. This revealed that although the pADAP-type plasmids do not appear to actively transfer genetic elements, there are numerous non-pADAP plasmids seemingly capable of still acquiring and transferring novel material. To demonstrate the complexity of bacterial plasmids, a synteny analysis for all plasmids assessed in this study was performed. The synteny blocks were converted to a binary absence/presence matrix to show regions shared with different plasmids (Figure 13-3). These regions however vary in size which is not reflected in the figure.

Full nucleotide sequence identity between all plasmids was obtained though alignment of all plasmids in this study using the MAFFT toolset (405). Results were converted into a nucleotide distance matrix, presented in Figure 13-4. These figures show that even when excluding the chromosomal differences between isolates, the differences in the plasmid genomes of closely related bacteria can be vast, and much effort is needed to model the exact functional impact this genomic disparity has on bioactivity.

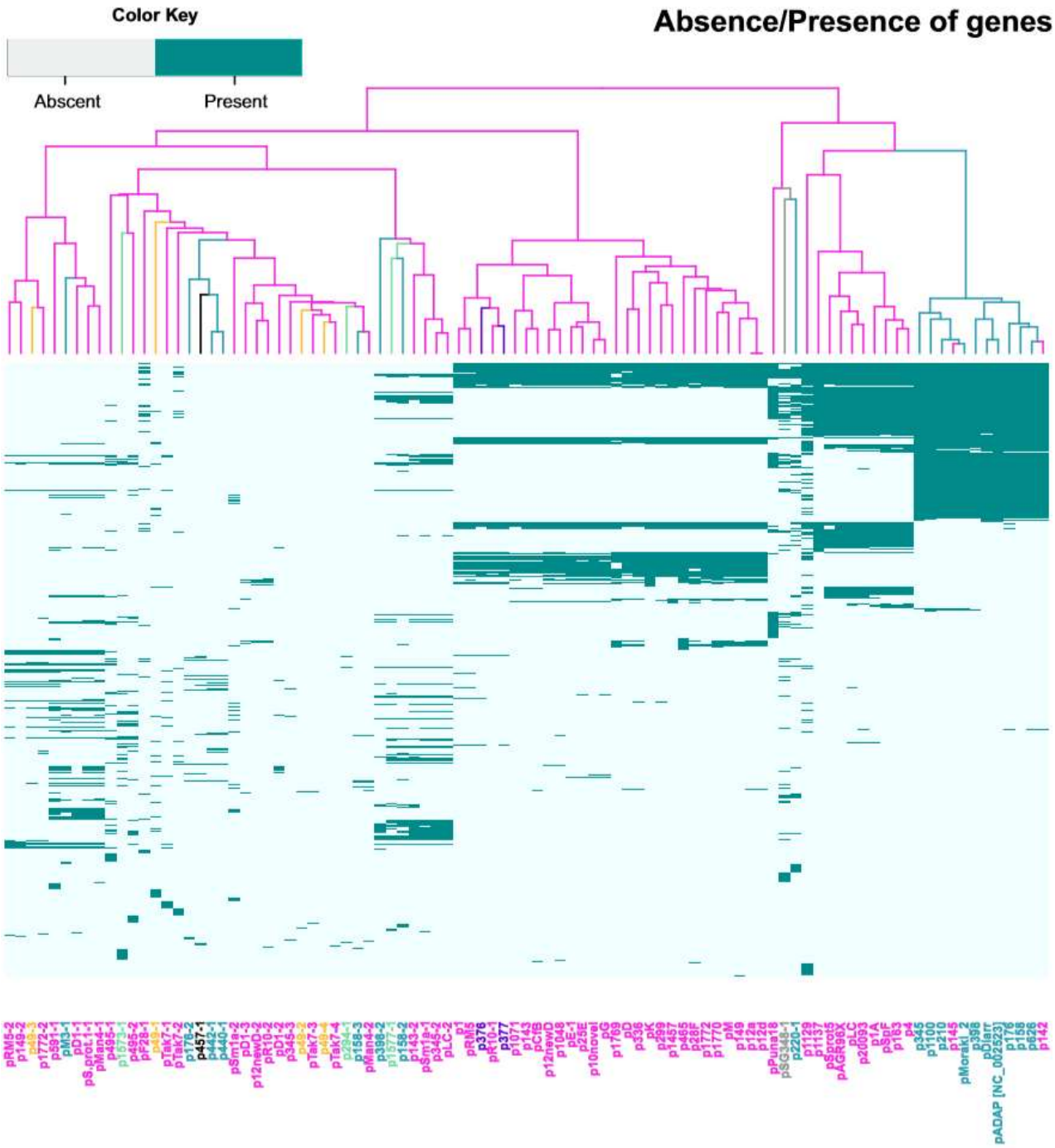


Figure 13-3: Absence/presence matrix of shared synteny blocks among several pADAP-type and non-pADAP plasmids. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens*, pink *S. proteamaculans*, grey *S. grimesii*, black *S. ficaria*, orange *Y. frederiksenii*. Analysis was performed using progressiveMauve (393). Full dataset including block sizes and location are provided in Supplementary Materials S.44.

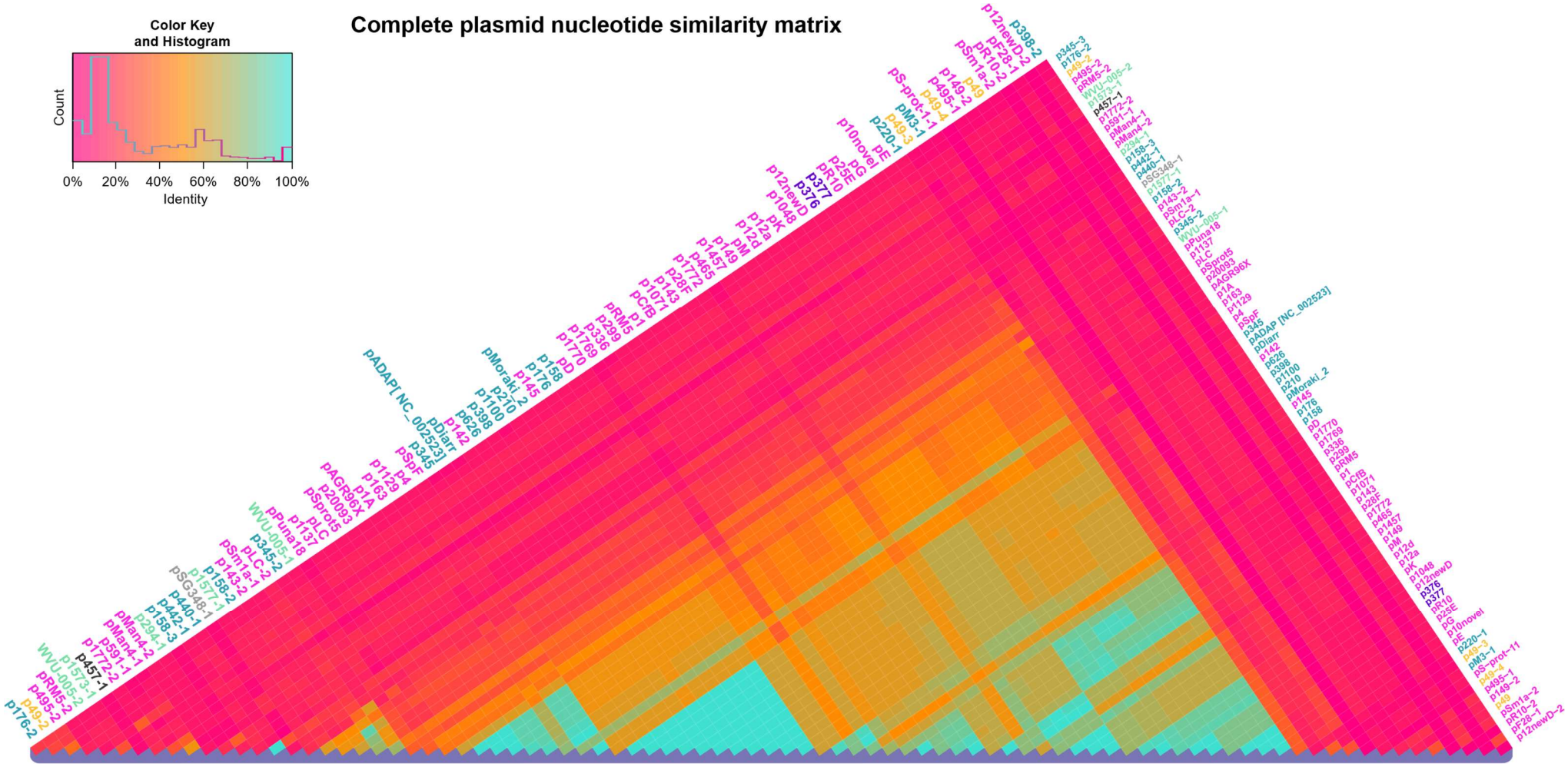


Figure 13-4: A DNA similarity matrix of the complete sequence of all plasmids sequenced in this study. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens*, pink *S. proteamaculans*, grey *S. grimesii*, black *S. ficaria*, orange *Y. frederiksenii*. Larger version of this image is presented in Supplementary Materials S.45.

Chapter 14 General discussion

14.1 **Conclusion and Discussion**

The pADAP plasmid (81) was first isolated from a *C. giveni* bioactive strain designated A1MO2 (77), resulting in the formation of two insecticidal products based on the phage resistant strain 626. The first product being Invade™ by Coated Seed NZ Ltd. (65, 78) and the second being Bioshield™ by BioStart™ (87), both garnering some commercial success. The pADAP plasmid was found to encode for two separate insecticidal elements, the Sep Toxin complex (161) and the Afp anti-feeding prophage particle (161). New isolates with bioactivity were believed to contain the same clonal pADAP plasmid until sequencing revealed pADAP elements having been horizontally transferred to a *Y. frederiksenii* 49 plasmid (90) and new Afp variants with different pathogenic phenotypes were found such as AfpX found in *S. proteamaculans* AGR96X (69). Based on these findings, this study commenced with the goal of sequencing a large number of inhouse isolates with different geographical, phenotypical and experimental features to analyse the genetic diversity of the pathogenic elements found on the plasmids and determine whether these elements were horizontally acquired, and to experimentally prove active transfer of these elements to other cells.

The data accumulated through the course of this study enabled further prediction of the origin of the pADAP plasmid family and its acquired regions. The *S. entomophila* related pADAP-type plasmids share a high percentage of similarity in both phenotype, gene absence/presence and nucleotide level, outliers such as the two pADAP carrying *S. proteamaculans* isolates 142 and 145 notwithstanding. The previously observed divergent band patterns in *S. proteamaculans* on Kado and Liu plasmid visualization gels (81), as well as DraI restriction profile gels (84, 94), suggested genetically diverse Sep and Afp PAIs. Through genomic sequencing and annotations, it has become clear the diversity presented on these plasmid visualization and restriction gels reflected a heavily diverse set of plasmids, both pADAP-types and non-pADAP types. These results are similar to other studies on comparative plasmidomics undertaken with *Rickettsia* (406), *Sulfolobus* (407), *Klebsiella pneumoniae* (408), *Listeria monocytogenes* (409) and in *Bacillus thuringiensis* and *Bacillus cereus* (63). In each of these studies, an array of dissimilar plasmid variants was identified, despite being hosted by closely related bacterial isolates belonging to one specific species or genus.

Plasmids have been shown to be a driver in the rapid adaptability of bacteria to new environmental factors due to low selection pressure that enables acquisition, deletion and optimization of encoded genes. One paper by Drake et al. (410) suggests the F plasmid found in *E. coli* has a five to 20 times higher mutation rate than the chromosome, which is estimated to be 1/300 base per genome per replication. This results in plasmids evolving quicker than their chromosomal counter parts, leading to chromosomally similar isolates having vastly differentiated plasmids.

The rapid evolution of these plasmids further stimulates genomic drift of the cell as it allows for speciation into niche micro milieus.

The diverse nature of pADAP-type plasmids might be a result of strains residing in a specific niche where pathogenicity is not under high selective pressure, for example, for isolates residing in areas void of grass grub, or alternatively, being associated with a yet to be determined host organism. Another possibility is that certain sub species of *Serratia* take on a role of "cheats", and profit from the pathogenicity of bioactive pADAP-type carrying *Serratia* species. Further to this, the hyper-virulent nature of some variants may limit their capacity to spread, where rapid infection and death of the host, limits the opportunity to spread, through reduced host contact. This may make convergence of all pADAP-type plasmids towards one efficient plasmid unlikely.

There is no observable incompatibility between pADAP-type plasmids and *Serratia* isolates, based on conjugation experiments undertaken in this study and previous studies undertaken by Grkovic et al. (85) and Glare et al (86). Therefore, it is likely that the genetic diversity of pADAP is a result of isolates having limited or no contact and/or occupying different niches.

This study identified several novel regions including a previously undescribed putative exotoxin effector present in several pADAP-type plasmids, a novel Afp variant designated as AfpA, found on the p1137 plasmid, a bacteriophage found inserted between the between *pilO* and *pilN* ORF in the conjugative region of p465, and several novel regions downstream of the Afp, AfpA and AfpX regions, one of which encodes a VapBC toxin-antitoxin system.

An attempt was also made to confirm whether the Sep and Afp pathogenicity associated islands (PAIs) are still mobile through mutating the plasmids with antibiotic cassettes and observing the loss of resistance, both *in vivo* and *in vitro*, but no loss was observed. Fitness experiments revealed that the A1MO2 pADAP carrying strain has fitness benefits over the pADAP heat-cured strain 5.6 in several growth conditions including nutritional stress as well as under MitC induced DNA damaging stress.

Based on all these findings, it can therefore be concluded that these pADAP plasmids are highly stable, and do not pose much of a burden on the host's fitness in the confines of the experimental design. Based on the similarity of the pADAP backbone from plasmids found in *S. entomophila*, *S. proteamaculans* and *S. marcescens*, combined with aforementioned experiments performed in this study and studies done by Glare et al. (86) and Grkovic et al. (85), it appears the pADAP type plasmids are easily transmittable to non-plasmid bearing *Serratia* cells that, once infected with the plasmid, are unable to easily rid themselves of it.

A major factor in the genetic diversity of the pADAP plasmid family is most likely the limited contact between pADAP-type plasmid carrying bacterium and their ability to interact with the greater *Serratia* supergenome. The *S. entomophila* and *S. proteamaculans* species that carry pADAP are tightly associated with the grass grub larvae, as bioassays on closely related species often show no bioactivity. It is also known that grass grubs, upon reaching the beetle stage, do not traverse long distances from their site of emergence, therefore limiting the distance that pADAP-type plasmid carrying *Serratia* bacteria can spread.

Another factor limiting spread is the death of infected larvae, reducing adult mediated transfer of the *Serratia* pathogen. Therefore it is likely that the majority of the genetic diversity, seen in the samples analyzed in this study, are a localised homogenization of the pADAP-type plasmid for that specific area, where interaction with non-pADAP carrying bacteria may increase the frequency of HGT of the plasmid.

As each region differentiated its pADAP and allows *Serratia* species to enter new niches, the interaction with these differenntiated pADAP is further diminished, resulting in distinct pADAP genotypes. The driving force between the pADAP disparency is mostly due to HGT facilitated natural selection, vertical transmission and point mutations, and less likely due to active HGT of large genomic islands and direct genetic recombination between pADAP elements.

The Figure 14-1 is provided to illustrate the previously outlined reasons behind the limited spread of the pADAP-type plasmids.

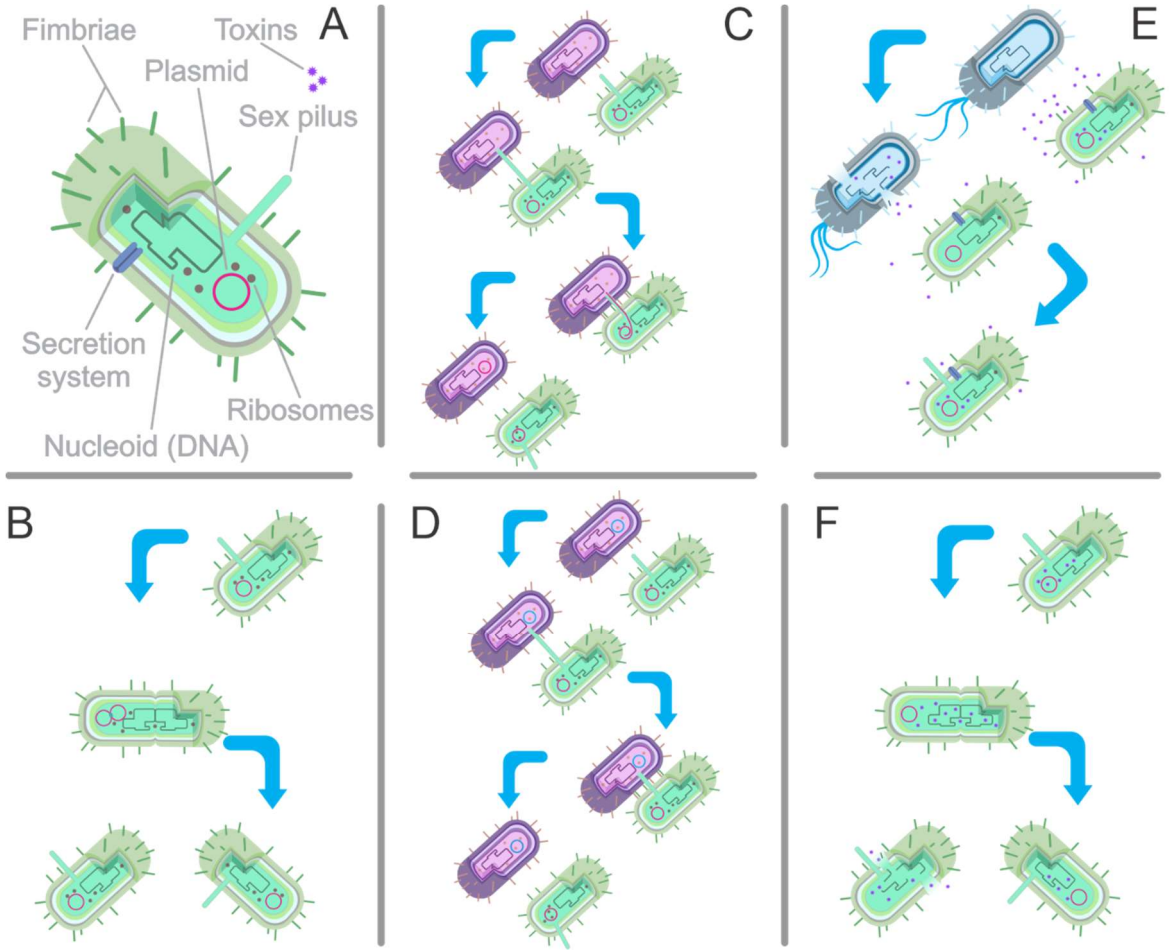


Figure 14-1: Examples of limitations of pADAP spread and convergence towards one pADAP genotype. A) a simplified schematic of the *Serratia* bacterium, B) Plasmids are single copy, as additional plasmids will result in dimerization of the two in a process called handcuffing, C) Plasmids can conjugate to other *Serratia* cells that experience no chromosomal anomalies that inhibit pADAP stability, D) Plasmids cannot conjugate to other *Serratia* cells already carrying a pADAP-type plasmid due to the reason mentioned in (B), E) Several plasmids carry TA operons that produce anti-microbial compounds, preventing interaction between plasmid bearing and plasmid free bacteria, F) Several plasmids carry PSKS TAs that produce toxic compounds, killing or debilitating progeny deprived of a pADAP replicon.

14.2 **Concluding remarks**

Although there are many plasmid-based factors involved with virulence, two of four of Koch's postulates revolve around the notion that disease associated effectors should always cause disease. However, none of the disease inducing isolates discussed in this study or in previous studies were shown to have a 100% success rate with a maximum dose challenge. This indicates that there is much more to learn about the disease states of the chronic Amber disease, and the hyper-pathogenicity induced by AfpX encoding plasmids. Several AfpX encoding pADAP-type plasmid carrying isolates also show that the hyper-pathogenicity observed in bioassays performed with the AGR96X type strain are unique to that strain, and not unique to the plasmid. Therefore, there must be additional chromosomal factors involved, potentially regulating AfpX production, or help with host infection in some unknown way. The work in understanding the chromosomal aspect of these diseases is currently being undertaken in another PhD study by Amy Vaughan.

It is also a known fact that bioassays performed on different days, different seasons or using grass grub larvae obtained from different regions, can occasionally have surprisingly different results, with one isolate showing as much as a 15.1% standard error rate based on triplicate bioassays. This suggests that the grass grubs themselves also have means of preventing disease, and it has been postulated that there might be genetic components involved, or even *C. giveni* sub species. Little work has been undertaken on this endemic species with deep rooted ties to New Zealand culture and agricultural industry and research in this field could identify weaknesses or resistance mechanisms present in grass grub larvae that could become useful in future population control.

Supplementary materials

The following documents are in support of statements made in previous chapters. Supplementary data can be downloaded from https://www.doi.org/10.25400/lincolnuninz.12736217

S.1 Sequenced samples

The following document is an excel sheet containing all the characteristics associated with the sequenced strains. This includes names and alternative names, locations, sequencing technologies, plasmid size based on the Kado and Liu megaplasmid visualization protocol (219) undertaken by Dodd et al (94), pathogenicity values based on maximum oral challenge bioassays (81) undertaken by Hurst [unpublished data] and this study, plasmid size based on assembly results (this study), the breakdown of the elements contained on the pADAP plasmid such as AFP, SEP and backbone (this study), classification of variant types, and any additional comments.

S.2 M9 minimal medium recipe

The following tables contain the recipe for M9 minimal medium

Table S-1: Recipe for M9 minimal medium.

M9 5x Stock	
64 g	Na2HPO4-7H2O
15 g	KH2PO4
2.5 g	NaCl
5.0 g	NH4Cl
Adjust to 1000 ml with distilled H2O	

M9 working solution	
700 ml	MilliQ dH2O
200 ml	M9 (5x) Stock
2 ml	1M MgSO4
20 ml	Casamino acid (20% w/v)
100 µl	1M CaCl2
Adjust to 1000ml with MilliQ dH2O	

S.3 Kado and Liu buffers

The following tables contain the buffer recipes used to perform a Kado and Liu megaplasmid visualization.

Table S-2: Recipes for the Lysis solution and E buffer used in Kado and Liu megaplasmid visualization.

Lysis solution	
0.121 g	Tris base
6 ml	SDS (10%)
0.82 ml	2M NaOH
Make up to 20 ml with MilliQ dH20	

E buffer (20x)	
0.968 g	Tris base
3.3 ml	3M NaAc
4 ml	0.5M EDTA
Make up to 100 ml with MilliQ dH20	
Adjust pH to 7.9 with glacial acetic acid	

E buffer working solution	
90 ml	MilliQ dH2O
10 ml	E buffer (20x)

S.4 PCR reagents and settings

The following tables contain the reagents used for a PCR reaction, and the temperature and time settings for a default PCR program

Table S-3: The buffers and reactants used for PCR reaction.

Reagent	**1 reaction**
Taq DNA Polymerase PCR Buffer (Invitrogen™)	5 µl (10X)
MgSO4 (Invitrogen™)	2 µl (20 mmol)
dNTPs (Invitrogen™)	1 µl (25 µmol)
Milliq dH_20	39.8 µl
Purified DNA template (~25 ng/µl)	1 µl
Forward and Reverse Primers	0.5 µl of stock (50 µM)
Platinum™ Taq DNA Polymerase (Invitrogen™)	0.2 µl

Table S-4: The touchdown PCR program used in this study. * temperatures and time differed slightly based on size of amplified product and primer set.

Step	Temperature	Time	Repeats
1	95	3 m	x1
2	95	30 s	
3	58*	30 s	x5
4	72	1 m*	
6	95	30 s	
7	55*	30 s	x30
8	72	1 m*	
10	4	∞	

S.5 Substrate list

The following table contains the antibiotics used in this study and their default concentrations.

Table S-5: Substrate list. Unless stated otherwise, these are the stock concentrations used.

Abbreviation	Full name	Stock concentration[1]
Cm90	Chloramphenicol	90 mg / ml
Km50	Kanamycin	50 mg / ml
Ap400	Ampicillin	400 mg / ml
Sp160	Spectinomycin	160 mg / ml
Gm15	Gentamicin	15 mg / ml
Tc30	Tetracycline	30 mg / ml
ALA50	5-aminolevulinic acid	50 mg / ml
X-gal50	5-bromo-4-chloro-3-indolyl--D-galactoside	50 mg / ml
Ara20	Arabinose	200 mg / ml

[1] 1µl of stock was used per 1 ml of medium

S.6 Primer sets

The following document is an excel sheet containing all major primer sets used in this project.

S.7 Restriction Digestion reagents

The following tables contain the reagents used for a restriction digestion reaction of DNA material

Table S-6: Restriction digestion reagents. *Buffer is restriction enzyme dependant.

PCR product	
10 µl	Buffer*
1 µl	Restriction enzyme
20 µl	DNA
69 µl	MilliQ dH2O

Vector DNA	
10 µl	Buffer*
1 µl	Restriction enzyme
4 µl	Vector DNA
85 µl	d mg water

S.8 Cloning constructs

The following document is an excel sheet containing all constructs used for cloning

S.9 Bioassay example

This picture shows grass grub used to test bioactivity of novel pADAP-type carrying isolates. Most grubs in the left column are visually healthy, apparent from the greyish hue and darkened gut, and have ingested 100% of the previously provided carrot cube. The top two cells of the middle column contain dead grubs. The right column contains several diseased grubs characterized by the yellowish appearance and the lighter coloured gut, caused by Sep induced gut clearance. Many grubs in the right tray have not fed due to Afp exposure. Bottom right larvae is near death.

Figure S-1: An example of a grass grub bioassay. In this figure there are three rows of 6 grubs, each column is fed with a different pathogenic bacterial strain. An empty column is added between sets to eliminate possibilities of cross contamination.

S.10 pADAP ORFs with new descriptions

The following table contains the reannotation of ORFs that were previously annotated with incorrect names or descriptions due to them not having any proper homologs in the NCBI database previously.

Table S-7: Updated functional description of previously described ORFs based on BlastX hits.

Original annotation	New annotation	Query Cover	E value	Per. Ident	Accession
anfa2	phage holin, lambda family [*Serratia entomophila*]	98%	2E-36	100%	WP_010895797.1
atoX	type II toxin-antitoxin system Phd/YefM family antitoxin [*Serratia entomophila*]	98%	5E-46	100%	WP_010895832.1
enp1	lysis protein [*Serratia entomophila*]	99%	7E-110	100%	WP_010895800.1
int1	IS3 family transposase [*Serratia entomophila*]	99%	4E-150	100%	WP_010895821.1
int2	tyrosine-type recombinase/integrase [*Serratia entomophila*]	99%	2E-173	100%	WP_010895831.1
mur1	lysozyme [*Serratia entomophila*]	99%	2E-99	100%	WP_010895802.1
parA	plasmid-partitioning protein SopA [*Serratia entomophila*]	99%	0	100%	WP_010895835.1
remS	N-6 DNA methylase [*Serratia entomophila*]	99%	0	100%	WP_010895837.1
spa3	NusG-type transcription antiterminator [*Serratia entomophila*]	99%	9.00E-116	98.75%	WP_010895762.1

S.11 Reannotation of sea ORFs and hypothetical proteins

The following table contains the reannotation of ORFs that were previously annotated as sea, due to them not having any proper homologs in the NCBI database.

This reannotation shows some additional hits although most appear to have some MGE-like functionality.

Table S-8: Reannotation of sea ORFs using BlastX.

ORF	New annotation	Query Cover	E value	Per. Ident	Accession
sea1	hypothetical protein				
sea2	hypothetical protein				
sea3	NusG-type transcription antiterminator [*Serratia entomophila*]	99%	6.00E-117	100.00%	WP_010895762.1
sea4	hypothetical protein				
sea5	replication regulatory protein RepA [*Serratia entomophila*]	98%	2.00E-59	100.00%	WP_010895764.1
sea6	DNA replication protein [*Klebsiella pneumoniae* subsp. Pneumoniae]	94%	7.00E-12	85.29%	PJR18670.1
sea7	Replication protein [*Serratia liquefaciens*]	85%	0	87.18%	WP_129983194.1
sea8	ATP-binding protein [*Serratia entomophila*]	99%	3.00E-80	100.00%	WP_010895767.1
sea9	ATP-binding protein [*Serratia entomophila*]	99%	0	100.00%	WP_010895767.1
sea10	SIR2 family protein [*Yersinia enterocolitica*]	99%	0	97.92%	WP_100273950.1
sea11	GIY-YIG nuclease family protein [*Serratia entomophila*]	99%	0	100.00%	WP_010895770.1
sea12	ATP-dependent helicase [*Enterobacter asburiae*]	98%	0	75.91%	WP_080972471.1
sea12'	DEAD/DEAH box helicase [*Escherichia coli*]	99%	5.00E-59	81.65%	SPW42256.1
sea13	class I SAM-dependent DNA methyltransferase [*Dickeya paradisiaca*]	96%	8.00E-105	80.00%	WP_015854718.1
sea14	ISKra4 family transposase [*Yersinia mollaretii*]	99%	4.00E-131	97.83%	WP_145559240.1
sea15	ISKra4 family transposase [*Hafnia alvei*]	99%	5.00E-145	86.96%	WP_103008330.1
sea16	hypothetical protein				
sea16'	hypothetical protein				
sea17	transcriptional regulator [*Enterobacter* sp. A8]	98%	8.00E-32	33.19%	WP_130626660.1
sea18	hypothetical protein				

sea19	EAL domain-containing protein [*Serratia entomophila*]	99%	0	100.00%	WP_010895790.1
sea20	hypothetical protein				
sea21	hypothetical protein				
sea22	Transposase [*Serratia fonticola* AU-AP2C]	51%	4.00E-12	80.56%	ERK06032.1
sea23	IS91 family transposase [*Vibrio vulnificus*]	100%	9.00E-15	81.40%	POC78681.1
sea24	ISNCY family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895822.1
sea25	ISNCY family transposase [*Serratia* sp. S4]	81%	1.00E-21	91.04%	WP_083906151.1
sea26	ISNCY family transposase [*Serratia entomophila*]	99%	8.00E-63	100.00%	WP_010895822.1
sea27	ISNCY family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895825.1
sea28	ISNCY family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895826.1
Sea29	nuclear transport factor 2 family protein [*Serratia* sp. BW106]	57%	7.00E-11	84.85%	WP_099062043.1
sea30	DUF2913 family protein [*Serratia entomophila*]	99%	1.00E-140	99.48%	WP_010895828.1
sea31	latent-transforming growth factor beta-binding protein 1-like [*Meriones unguiculatus*]	67%	0.84	32.18%	XP_021508259.1
sea32	5’/3’-nucleotidase SurE [*Halovivax asiaticus*]	34%	5.2	26.67%	WP_007698437.1
sea33	glycosyl transferase family 1 [*Crocinitomicaceae bacterium*]	25%	1.4	38.03%	MBG27488.1
sea34	hypothetical protein				
sea35	StbA family protein [*Serratia liquefaciens*]	99%	7.00E-103	98.06%	WP_128865566.1
sea36	helix-turn-helix transcriptional regulator [*Serratia entomophila*]	99%	2.00E-119	100.00%	WP_010895843.1
Sea37	MFS transporter [*Brevibacterium iodinum*]	76%	0.59	32.69%	WP_101546461.1
sea38	aerotaxis sensor receptor, flavoprotein [*Escherichia coli*]	70%	2.00E-11	41.24%	WP_097331443.1
sea39	hypothetical protein				
Sea40	mRNA interferase HigB (modular protein) [*Serratia symbiotica*]	99%	8.00E-65	95.05%	CDS57329.1
sea41	type II toxin-antitoxin system HigA family antitoxin [*Salmonella enterica*]	99%	4.00E-62	63.27%	EAB5051308.1
sea42	hypothetical protein				
Sea43	hypothetical protein				
sea44	hypothetical protein				
sea45	autotransporter adhesin Ag43 [*Escherichia coli*]	98%	3.00E-07	27.27%	WP_001825652.1
sea46	hypothetical protein				
Se3	IS66 family transposase [*Serratia entomophila*]	99%	0	100.00%	WP_010895794.1
se2	IS66 family insertion sequence element accessory protein TnpB [*Serratia entomophila*]	98%	1.00E-63	100.00%	WP_010895795.1
se1	IS66 family insertion sequence hypothetical protein [Serratia entomophila]	99%	5.00E-83	100.00%	WP_010895793.1

Hypothetical protein	DUF2514 family protein [*Serratia entomophila*]	99%	2.00E-101	100.00%	WP_010895740.1
Hypothetical protein	lysozyme [*Serratia entomophila*]	99%	2.00E-101	100.00%	WP_010895739.1
Hypothetical protein	hypothetical protein				
Hypothetical protein	phage IX holin family protein [*Serratia entomophila*]	98%	2.00E-57	100.00%	WP_010895738.1

S.12 Plasmid copy number from coverage plot

The following documents are custom Bash shell scripts containing the steps used to generate the coverage per contig as well as the custom R script used to convert the output to an understandable plot. Both scripts contain hardcoded paths and data specific naming conventions and should therefore only serve as examples.

https://github.com/IamIamI/pADAP_project/tree/master/Plasmid_copy_from_coverage

S.13 RepA orthologs used in this study and subsequent alignment

The following table shows the reference replication proteins used for the RepA comparison.

Table S-9: All orthologous RepA proteins that were used in the RepA comparison.

Species		Accession	Association	Reference
Erwingella	*americana*	[WP_034796133]	Opportunistic	Brenner et al. (281)
Erwinia	*amylovora*	[WP_015056247]	Plant	Rhim et al. (278), Kube et al. [251]
Edwardsiella	*tarda*	[WP_068872131]	Fish	Wang et al (288)
Enterobacter	*cloacae*	[WP_040021347]	Mammals	Lui et al. (286)
Enterobacter	*hormaechei*	[WP_058658352]	Mammals	Townsend et al. (287)
Erwinia	*billingiae*	[WP_053140920]	Plant	Kube et al. (279)
Erwinia	*persicina*	[WP_062749059]	Plant	Zhang et al. (277)
Erwinia	*pyrifoliae*	[WP_011078049]	Plant	Rhim et al. (278)
Escherichia	*coli*	[WP_085447277]	Mammals	Souza et al. (285)
Nissabacter	*archeti*	[WP_072928254]	Mammals	Mlaga et al. (280)
Rahnella	*aquatilis*	[WP_014341718]	Opportunistic	Brady et al. (282)
Rahnella	*bruchi*	[WP_120508591]	Insect	Brady et al. (282)
Rahnella	*victoriana*	[WP_095925420]	Plant	Brady et al. (282)
Rahnella	*woolbedingensis*	[WP_120134817]	Plant	Brady et al. (282), Brenner et al. (281)
Salmonella	*enterica*	[EAM9793237.1]	Mammals, Bird	Rabsch et al. (289)
Serratia	*grimesii*	[ADR74383.1]	Opportunistic	Mahlen et al. (290)
Serratia	*marcescens*	[WP_060560535.1]	Mammals	Mahlen et al. (290)
Serratia	*entomophila pADAP*	[NC_002523]	Insect	Grimont et al. (18)
Serratia	*proteamaculans pPuna18*		Plant	* This study
Serratia	*proteamaculans p1137*		Insect	* This study
Serratia	*proteamaculans*	[NC_002523]	Insect	Hurst et al. (69)
Yersinia	*kristensenii*	[WP_050116464]	Mammals	Robins-Browne et al. (284)
Yersinia	*pseudotuberculosis*	[WP_011191397]	Opportunistic	Martínez-Chavarría et al. (283)

The following documents are MEG files containing an alignment between several pADAP RepA sequences and their closest RepA orthologs in the NCBI database and the corresponding phylogenetic ML tree in nexus format, generated from this alignment using PhyML.

S.14 Iteron comparison reference strains and alignment

The following table shows reference replication genes used for the iteron comparison. pADAP was used as the representative for all pADAP-type plasmids obtained from this study as they all exhibited a highly similar iteron pattern as well as nucleotide conservation for the region.

Table S-10: All homologous *repA* plasmids that were used in the iteron region comparison. [1] Isolate were obtained from France. [2] Isolate was obtained from Argentina.

Accession number	Description	Strain	Plasmid
AY123045	*Erwinia pyrifoliae*	Ep1/96	pEP36
CP009710	*Yersinia pseudotuberculosis*	IP 32953	pYAC_2
CP022727	*Erwinia persicina*	B64	pEP2
CP032297	*Rahnella aquatilis*	ZF7	pRAZF7
HG793097	*Erwinia amylovora*	LA636	LA636P1
NC_002523	*Serratia entomophila*	A1MO2	pADAP
-	*Serratia proteamaculans*	F28	pF28.p1
-	*Serratia entomophila*[1]	220	p220.p1
-	*Serratia grimesii*[1]	348	p348.p1
-	*Serratia proteamaculans*[2]	495	p495.p1

The following documents are a MEG file containing an alignment between the iteron regions of all pADAP plasmid backbones extracted from the 53 pADAP plasmids and a .gb file containing the regions with full feature annotation.

S.15 pADAP backbone start

The following document is a MEG file containing an alignment of the region upstream of all pADAP *repA* sequences. This shows a clear point where sequence identity drops substantially and proposed to be the start of the pADAP backbone.

S.16 pADAP plasmid backbone alignment

The following documents are MEG files containing an alignment between all pADAP plasmid backbones extracted from the 53 pADAP plasmids and a .gb file containing regions with full feature annotation. Alignment was produced using Clustal Omega.

S.17 Deviating ORFs of pADAP-type plasmid backbones

The following file is a manually curated excel file, containing observed dissimilarities between gene annotations on the pADAP backbone. The file contains numerous types of aberrant gene annotations compared to the reference pADAP [NC_002523]. This analysis does not include hypothetical gene annotations with no significant tBlastX hit, not does it include the two major backbone regions discussed in in Chapter 4.3, as these are too genetically diverse to manually describe.

S.18 Distance matrix files

The following files are csv files containing distance matrices for different regions. These matrices were produces by Geneious and are based on sequence similarity on a nucleotide level.

S.19 Distance matrix processing R-script

The following file is an R script. The script should be able to run standalone. It required a distance matrix such as produced by Geneious. Example of the format can be found in Supplementary Materials S.43. The script will produce 2 prompts;

- the first is a windows explorer popup in which the user can pick the gene_presence_absence.csv file
- Second prompt is to pick a location and name for the sequence similarity matrix plot

The plot produced by this script was manually edited using CorelDraw. The triangular matrix was flipper ~220°, the labels were manually stylized, a purple bottom bar was added to make it look level, other minor tweaks were made. These changes can also easily be undertaken using a free graphical tool such as InkScape.

https://github.com/IamIamI/pADAP_project/tree/master/Distance_Matrices

S.20 Demarcation region alignment

The following documents are MEG files containing an alignment between all transposon islands extracted from the 53 pADAP plasmids and a .gb file containing regions with full feature annotation. Alignment was produced using Clustal Omega.

S.21 Sef region alignment

The following documents are a MEG file containing an alignment between all Sef / Sep regions of plasmid backbones extracted from both pADAP and non-pADAP plasmids and a .gb file containing regions with full feature annotation. Alignment was produced using Clustal Omega.

S.22 SEF expression in AAE072CE Fim null E. coli strain vs AAE072CE type strain

The following figure shows several additional TEM pictures of AAE072CE + pARA_SEF_expression vs. AAE072CE + pARA. This figure illustrates that the SEF region is encoding a functional fimbria.

Figure S-2: A comparison between the fimbriae expressed AAE072CE Fim null E. coli strains + pARA_SEF_Cm and the non-fimbriae expressing AAE072CE type strain.

S.23 Sep / Spp region alignment

The following documents are a MEG file containing an alignment between all Sep regions of plasmid backbones extracted from both pADAP and non-pADAP plasmids and a .gb file containing regions with full feature annotation. Alignment was produced using Clustal Omega.

S.24 Afp ORF function and homologs

The following table contains the description of Afp ORFs.

Table S-11: The role of individual Afp ORFs in the formation and functionality of Afp particles.

Original annotation	Function
afp1, 5;	Afp1 is a tube protein, Afp5 the tube initiator protein (202), high amino acid similarity with gp19 protein derived from T4 phage tail tube which is a rigid tube consisting of several gp19 and used to perforate the target cell (201).
Afp2 – 4;	Sheath proteins (202), highly homologous to the sheath domains of bacteriophages T4 and P2 (198).
Afp6 – 7;	Needle and tube proteins (202)
afp8;	Needle base (202), high similarity with a T4 phage gp5-gp27 puncturing device (194).
Afp9;	Sheath initiation protein (202), homologous to baseplate-related proteins produced by *Photorhabdus luminescens* PVC (198).
Afp10 – 12;	Needle tip, and baseplate wedges (202)
afp13;	Tail fibres (202), Composed of a three conserved amino acid repeats that show high amino acid similarity with adenovirus tail fibres (198).
Afp14;	Tape measure protein, a protein that defines phage length (203).
Afp15;	ATPase domain containing protein (198). Homologous to a predicted ATPase protein called ClpV [169]. ClpV is part of protein secretion system in Pseudomonas aeruginosa, its function being to release energy needed for the secretion (411).
Afp16;	Terminates elongation of Afp tube and initiates Afp sheath polymerization (201).
Afp17 – 18;	Putative toxins, both have a low G+C% rating, consistent with genes acquired through HGT (198). When afp18 was not present, there was no cessation of feeding in grass grub, suggesting that it is the toxic component of the Afp. The proteins Afp17 and Afp18 as independent entities, do not result in cessation of feeding, suggesting that the components only work when internalized within the target cell, mediated through the product formed from Afp1-16 (201).

S.25 Afp region alignment

The following documents are a MEG file containing an alignment between all Afp regions of plasmid backbones extracted from pADAP plasmids and a .gb file containing regions with full feature annotation. Alignment was produced using Clustal Omega.

S.26 Non-grass grub species

The following figure shows several non-grass grub species used to test additional bioactivity of novel pADAP-type carrying isolates. Red headed cockchafer is noted to have a “reddish” head and is twice the size of a grass grub. Odontria are larger than grass grub and the head is slightly browner. Tasmania grass grub have a slightly smaller body than grass grub and have a dark purple head. Manuka beetle are half the size as a third instar grass grub and can be easily confused for first instar grass grubs.

Figure S-3: Examples of third instar larvae of several New Zealand endemic beetle species. These species are compared to larvae of the grass grub (*Costalytra giveni*) seen at the bottom of each picture.

S.27 Amino acid alignment of the Afp13 and Afp18 from p1137 to that of pADAP

The following figure shows the amino acid alignment of Afp13 from p1137 to that of pADAP.

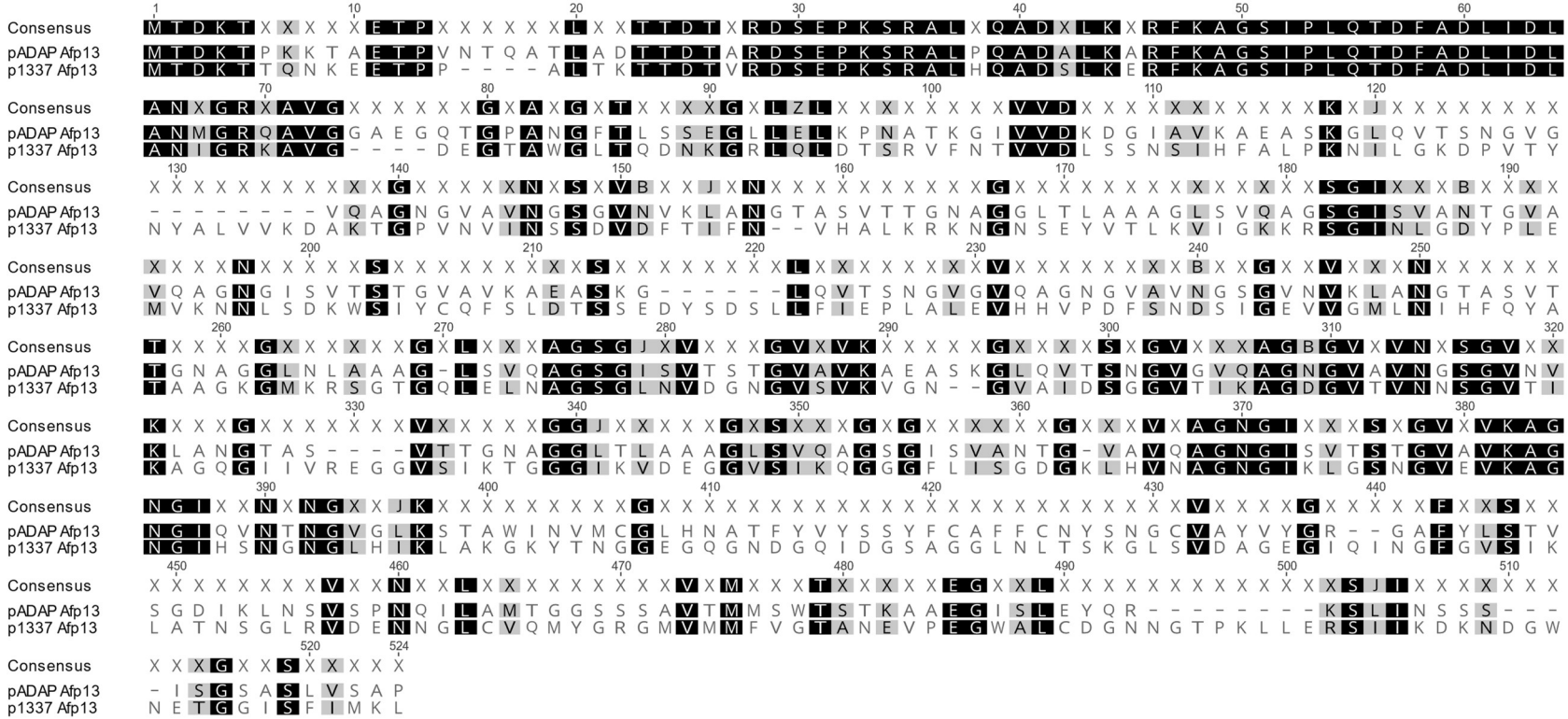


Figure S-4: Alignment of pADAP Afp13 and p1137 Afp13A.

The following figure shows the amino acid alignment of Afp18 from p1137 to that of pADAP.

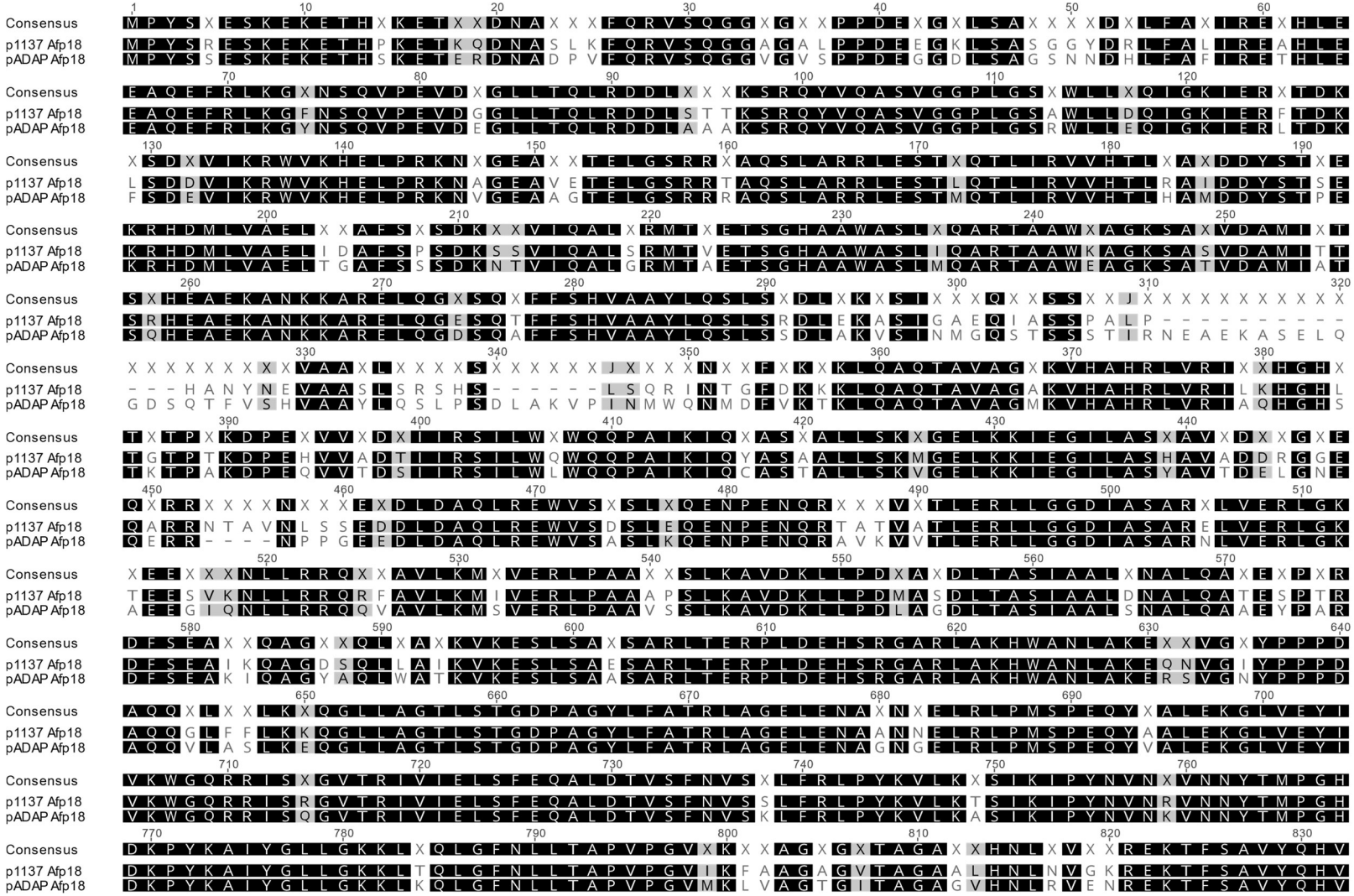
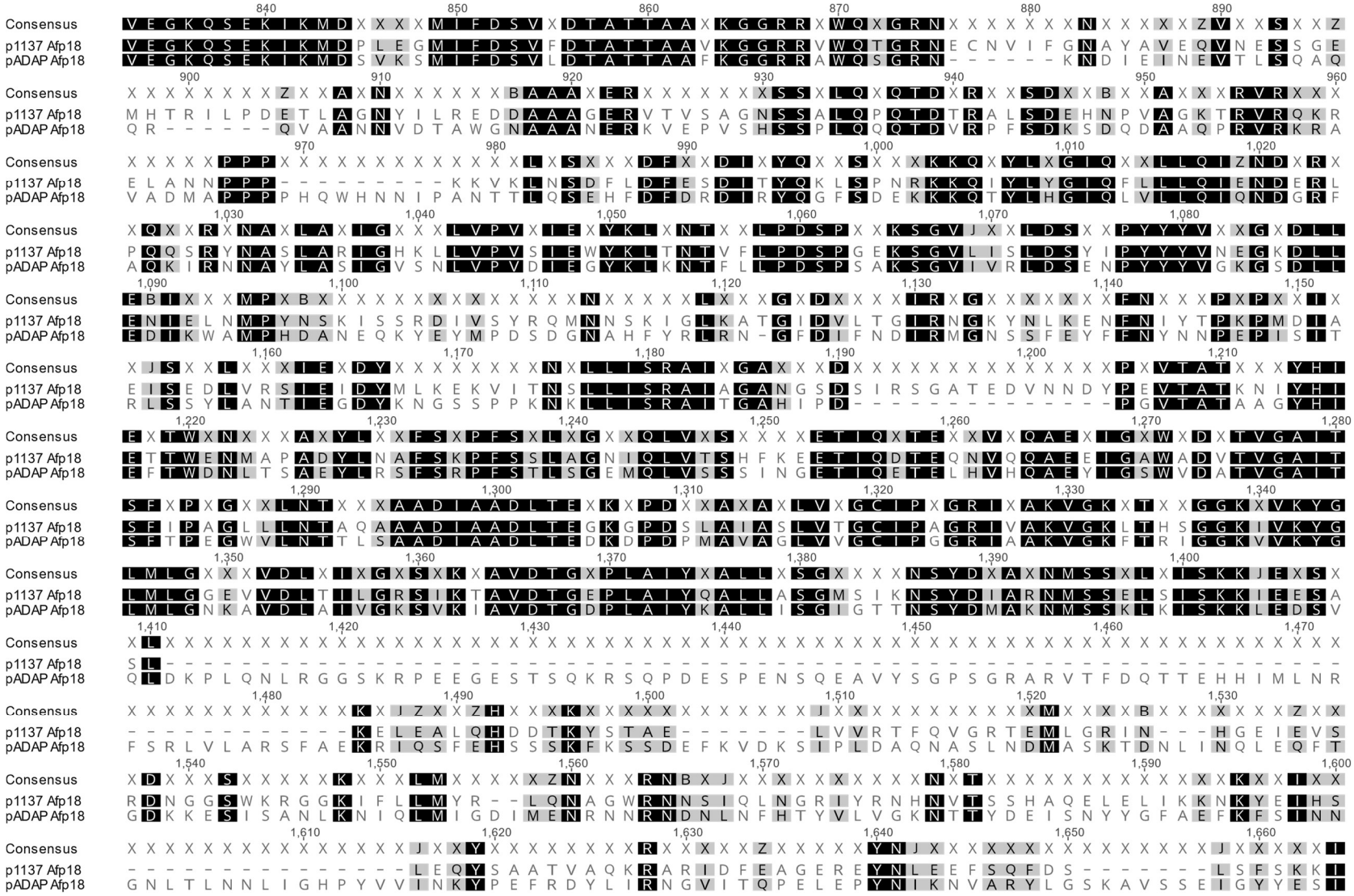
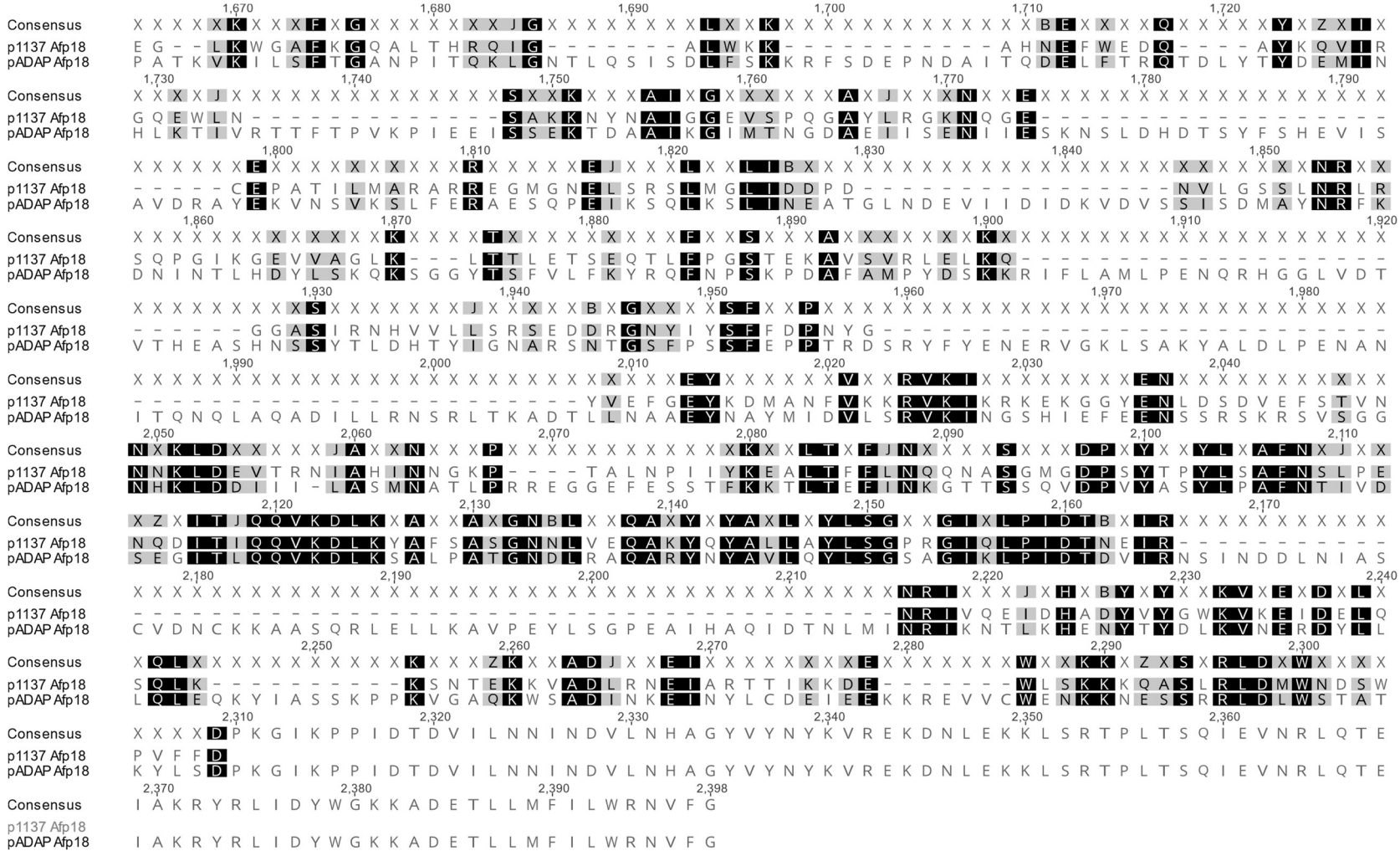


Figure S-5: Alignment of p1137 Afp18A and pADAP Afp18.


Consensus
p1137 Afp18
pADAP Afp18


Consensus
p1137 Afp18
pADAP Afp18

S.28 Annotation of p149 novel region

The following table contains the annotations of ORFs that were present in the Novel region of p149 and similar pADAP-type plasmids.

Table S-12: Description of ORFs found in the novel region of reference isolate 149.

Orf	Size	Annotation	Query Cover	E value	Identity	Accession
1	64	hypothetical protein [*Salmonella enterica*]	95%	9.00E-27	72.13%	EBB0713068.1
2	52	tyrosine-type recombinase/integrase [*Serratia proteamaculans*]	98%	4.00E-22	88.24%	WP_129937181.1
3	179	MULTISPECIES: stationary phase growth adaptation protein [*Serratia*]	99%	2.00E-125	96.63%	WP_048796374.1
4	99	No hits				
5	114	single-stranded DNA-binding protein [*Serratia plymuthica*]	96%	2.00E-62	81.82%	WP_037429931.1
6	281	ParA family protein [*Serratia marcescens*]	99%	0	93.57%	WP_038883455.1
7	378	ParB/RepB/Spo0J family partition protein [*Gibbsiella quercinecans*]	99%	0	71.99%	WP_121554069.1
8	146	hypothetical protein [*Gibbsiella quercinecans*]	99%	9.00E-80	78.62%	WP_121554070.1
9	108	hypothetical protein [*Serratia plymuthica*]	99%	7.00E-61	83.18%	WP_100220740.1
10	408	hypothetical protein [*Serratia plymuthica*]	99%	0	86.88%	WP_004953880.1
11	347	IS481 family transposase [*Salmonella enterica* subsp. enterica serovar Oranienburg]	99%	0	98.55%	EBZ5718334.1
12	180	phage tail protein [*Pectobacterium parmentieri*]	82%	4.00E-18	36.77%	WP_121296198.1
13	58	IS110 family transposase [*Cedecea* sp. NFIX57]	94%	2.00E-16	68.97%	WP_139827421.1
14	110	unnamed protein product [Pseudo-*nitzschia multistriata*]	64%	8.9	30.99%	VEU38783.1
15	193	MULTISPECIES: response regulator transcription factor [*Serratia*]	97%	2.00E-20	35.96%	WP_122080126.1
16	173	fimbrial protein [*Serratia plymuthica*]	98%	3.00E-75	69.44%	WP_126527321.1
17	862	outer membrane usher protein [*Serratia fonticola*]	96%	0	67.27%	WP_141133248.1
18	251	fimbria/pilus periplasmic chaperone [*Serratia fonticola*]	91%	1.00E-122	72.93%	WP_147281001.1
19	188	type 1 fimbrial protein [*Serratia nematodiphila*]	94%	1.00E-65	57.30%	WP_139792520.1
20	165	exotoxin [*Serratia nematodiphila*]	98%	2.00E-35	44.51%	OQV36223.1
21	163	fimbrial protein [*Serratia quinivorans*]	94%	1.00E-46	50.32%	WP_112350010.1
22	184	TPA: exotoxin [*Serratia* sp.]	82%	3.00E-55	57.62%	HCV64546.1
23	276	adhesin [*Photorhabdus* sp. MEX20-17]	83%	7.00E-45	36.86%	WP_132354722.1
24	517	hypothetical protein [*Pantoea* sp. OXWO6B1]	51%	2.00E-50	40.60%	WP_063877193.1
25	106	IS1400 transposase B [*Yersinia enterocolitica*]	99%	7.00E-70	99.05%	CQH32055.1

26	116	nuclear transport factor 2 family protein [*Serratia proteamaculans*]	50%	2.00E-33	98.28%	WP_129934637.1
27	453	ISNCY family transposase [*Erwinia* sp. OLMDLW33]	98%	0	94.64%	WP_099753821.1
28	161	antitermination protein NusG [*Yersinia* sp. IP36721]	98%	3.00E-48	49.37%	WP_120806842.1
29	98	RHS repeat-associated core domain-containing protein [*Yersinia* sp. IP36721]	93%	2.00E-36	69.57%	WP_120806841.1
30	1459	RHS repeat-associated core domain-containing protein [*Yersinia entomophaga*]	99%	0	58.98%	WP_064516743.1
31	172	MULTISPECIES: hypothetical protein [*Yersinia*]	98%	5.00E-51	47.75%	WP_064516741.1
32	318	MULTISPECIES: hypothetical protein [*Yersinia*]	99%	1.00E-141	68.32%	WP_064516739.1
33	418	MULTISPECIES: beta-ketoacyl-[acyl-carrier-protein] synthase family protein [*Yersinia*]	99%	0	75.54%	WP_064516737.1
34	263	MaoC family dehydratase [*Chromobacterium violaceum*]	99%	2.00E-43	33.21%	WP_011135094.1
35	86	MULTISPECIES: acyl carrier protein [*Yersinia*]	98%	6.00E-26	57.65%	WP_064516733.1
36	95	SppB [*Serratia proteamaculans*]	91%	1.00E-15	96.49%	ACZ05620.1
37	281	SppA [*Serratia proteamaculans*]	99%	3.00E-177	99.29%	ACZ05627.1

S.29 RUF region alignment

The following documents are a MEG file containing an alignment between all Novel regions of plasmid backbones extracted from both pADAP and non-pADAP plasmids and a .gb file containing regions with full feature annotation. Alignment was produced using Clustal Omega.

S.30 pPuna18 novel ORF annotations

The following document is an excel file containing the top functional descriptive BlastX hits for the ORFs present in the novel region of pPuna18.

S.31 Novel Bacteriophage ORF annotations, feature map and alignment

The files are an excel file containing the top functional descriptive BlastX hits for the ORFs present in the novel bacteriophage region. A MEG file containing an alignment of the bacteriophage region found in p465 and p591, and a .gb file containing the regions with full feature annotations.

S.32 Novel WVU-005-1 ORF annotations

The following document is an excel file containing the top functional descriptive BlastX hits for the ORFs present in the novel region of WVU-005-1.

S.33 Plasmid stability experiment

The following files are excel files containing the results of two large plasmid stability experiments and an archive of the raw data obtained from the flow cytometry experiment. The first file contains the results of a plate experiments undertaken with antibiotic markers tagged plasmid mutants, no plasmid loss was observed. The second file contains the results obtained from an analysis undertaken in FlowJo on a flow cytometry experiment undertaken with GFPmut3 tagged plasmid mutants. The final file is an archive of the raw flow cytometry output for each sample.

S.34 Growth curve experiment

The following table is a representation of the dilution scheme used for the cell culture plates that were to be analysed in the growth curve experiment.

Table S-13: A table showing the dilution ranges that were analysed for each growth condition.

	1	2	3	4	5	6	7	8	9	10	11	12
A	A1MO2	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
B	A1MO2	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
C	A1MO2	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
D	A1MO2	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
E	5.6	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
F	5.6	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
G	5.6	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank
H	5.6	1:10	1:10^2	1:10^3	1:10^4	1:10^5	1:10^6	1:10^7	1:10^8	1:10^9	1:10^10	Blank

The following files are excel files containing the results of three growth curve experiments. Each file contains the raw plate reader data, the blank corrected data, the average and the blank corrected average data for all quadruplicate replicates for both the A1MO2 plasmid bearing and the 5.6 plasmid cured samples obtained from the SPECTROstar Nano MARS data analysis software (223). The files are separated by medium as indicated in the file title; LB, M9 and LB + MitC.

S.35 Growth curve R-script

The following file is a R script. The purpose of this script is to convert the raw plate reader data obtained from SPECTROstar Nano MARS (223). This script contains hardcoded paths and data specific naming conventions and should therefore only serve as examples.

https://github.com/IamIamI/pADAP_project/tree/master/growth_curve_plots

S.36 Roary raw absence presence matrix

The following file is a csv file produced by Roary using the .gff files produced by the Prokka annotation program, for 53 pADAP-type plasmids. The cut off for calling orthologous groups was set at 90%.

S.37 Roary processing R-script

The following file is a R script. The script should be able to run standalone. It will produce 4 prompts;

- the first is a windows explorer popup in which the user can pick the gene_presence_absence.csv file
- Second prompt is to pick a location and name for a small plot of the absence/presence matrix without any gene labels
- Third prompt is to pick a location and name for a large plot of the absence/presence matrix that contains all the gene labels
- Fourth prompt is to pick a location and name for combined plot of several statistics such as total number of genes per genome, number of unique genes, number of core genes (based on the definition that 99% of the genomes have this gene), and several other stats.

https://github.com/IamIamI/pADAP_project/tree/master/Roary_stats

S.38 Roary plot, long format with readable y-axis labels

This plot shows the Roary clustering with an addition of readable y-axis labels. This helps with identifying the actual type of genes present in each orthologous group.

S.39 SCOARY output

The following files are csv file produced by SCOARY. The first file is the traits_table file needed for SCOARY to analyses binary traits in order to analyse whether there is a significant gene associated with that specific trait. All other files are output files produced by SCOARY that contain the genes that are potentially associated with the traits based on the Roary absence presence matrix.

S.40 Plasmid size and gene length distribution

The following document is a csv file containing the plasmid sizes on the first row, followed by a list of gene sizes. These values are raw values obtained by extracting the start and end of every CDS in a .gff file, and then calculating the size from that. This excel sheet was generate for use in R and is not formatted in a way usable to the reader.

S.41 Nucleotide bias analysis whole pADAP plasmids

The following documents are an excel sheets containing the results of a nucleotide biased analysis performed on 53 pADAP-type plasmids. This document shows the specific nucleotide preference that the CDS have. There is a discernible difference between the nucleotide distribution of the *S. entomophila* and *S. liquefaciens* versus that of the *S. proteamaculans* plasmids. Although there is an observable difference in distribution, the sample size of *S. entomophila* and *S. liquefaciens* strains is substantially smaller than that of the *S. proteamaculans*. The *S. proteamaculans* group also contains several outlier plasmids with unique features, with no homologous plasmid found in the *S. entomophila* and *S. liquefaciens* groups. The purpose of the script is to convert the nucleotide bias analysis tables produced by GCUA into a whisker plot. This script contains hardcoded paths and data specific naming conventions and should therefore only serve as examples.

S.42 CDS codon usage analysis whole pADAP plasmids

The following document is a excel sheet containing the results of a codon usage analysis performed on 53 pADAP-type plasmids. This document shows the specific codon preference that the CDS have for transcription of amino acids. No discernible difference is observed between plasmids residing in different chromosomal backgrounds, nor are there any observable differences between plasmids of different genotypes despite them having different genes or genetic islands. The purpose of this script is to convert a the CDS codon usage analysis tables produced by GCUA into a heatmap. This script contains hardcoded paths and data specific naming conventions and should therefore only serve as examples.

S.43 pADAP core backbone gene alignments and ML trees

The following documents are MEG file containing alignments for 13 core backbone genes shared by all 53 pADAP-type plasmids. Alignment was produced using Clustal Omega. Additional files are nexus files containing phylogenetic ML tree file for 13 core backbone genes shared by all 53 pADAP-type plasmids. Trees were generated using PhyML. A consensus Tree was generated using SumTrees from the DendroPy library.

S.44 High resolution summary figures

The following PDF shows the final summary figures in higher resolution for readability purposes

S.45 Complete plasmid dataset synteny analysis

The following PDF and excel file show shared synteny blocks between all plasmids analysed in this study. This serves only an anecdotal function, as little meaning can be derived from such a vast amount of dissimilarity between all plasmids. The excel file shows the regions and sizes of the synteny blocks per plasmid. PDFs shows the distance matrix based on full alignments of all plasmids analysed in this dataset as well as Roary analysis. This serves only an anecdotal function, as little meaning can be derived from such a vast amount of dissimilarity between all plasmids. Blue labelled strain names are *S. entomophila*, green *S. marcescens*, purple *S. liquefaciens*, pink *S. proteamaculans*, grey *S. grimesii*, black *S. ficaria*, orange *Y. frederiksenii*.

References

1. Coca-Abia MM, Romero-Samper J. Establishment of the identity of *Costelytra zealandica* (White 1846) (Coleoptera: Scarabeidae: Melolonthinae) a species commonly known as the New Zealand grass grub. New Zealand Entomologist 2016; 39: 129-146.
2. Richardson J, Gray JE, Ross JC, . E, . T. The zoology of the voyage of the H.M.S. Erebus & Terror, under the command of Captain Sir James Clark Ross, during the years 1839 to 1843. By authority of the Lords Commissioners of the Admiralty. Edited by John Richardson and John Edward Gray. London: E. W. Janson; 1844.
3. Given BB. A revision of the Melolonthinae of New Zealand. Pt. 1: the adult beetles. New Zealand Department of Scientific and Industrial Research Bulletin 1952; 102: 1-137.
4. Jackson TA, Townsend RJ, Nelson TL, Richards NK, Glare TR. Estimating amber disease in grass grub populations by visual assessment and DNA colony blot analysis. Proceedings of the New Zealand Plant Protection Conference 1997; 50: 165-168.
5. Townsend RJ. Grass grub lifecycle and control. https://wwwfarorgnz 2002; Arable Updates.
6. Bain J. Melolonthine beetles in forests (Coleoptera: Scarabaeidae: Melolonthinae). Grass grub and other chafers. Melolonthine beetles in forests (Coleoptera: Scarabaeidae: Melolonthinae) Grass grub and other chafers 1980.
7. Wightman JA. Rearing *Costelytra zealandica* (Coleoptera: Scarabaeidae). New Zealand Journal of Zoology 1974; 1: 217-223.
8. Perry GLW, Wheeler AB, Wood JR, Wilmshurst JM. A high-precision chronology for the rapid extinction of New Zealand moa (Aves, Dinornithiformes). Quaternary Science Reviews 2014; 105: 126-135.
9. McWethy DB, Whitlock C, Wilmshurst JM, McGlone MS, Fromont M, Li X, et al. Rapid landscape transformation in South Island, New Zealand, following initial Polynesian settlement. Proceedings of the National Academy of Sciences 2010; 107: 21343.
10. McWethy DB, Wilmshurst JM, Whitlock C, Wood JR, McGlone MS. A high-resolution chronology of rapid forest transitions following Polynesian arrival in New Zealand. PLoS ONE 2014; 9: e111328.
11. Levy EB. The conversion of rain forest to grassland in New Zealand. Tuatara 1949; 2: p37-51.
12. Lefort MC, Boyer S, Vereijssen J, Sprague R, Glare TR, Worner SP. Preference of a native beetle for "exoticism," characteristics that contribute to invasive success of *Costelytra zealandica* (Scarabaeidae: Melolonthinae). PeerJ 2015; 3: e1454.
13. East R, King PD, Watson RN. Population studies of grass grub (*Costelytra zealandica*) and black beetle (*Heteronychus arator*) (Coleoptera: Scarabaeidae). New Zealand Journal of Ecology 1981; 4: 56-64.
14. East R, Kain WM, Douglas JA. The effect of grass grub on the herbage production of different pasture species in the pumice country. Proceedings of the New Zealand Grassland Association 1980; 41: 105-115.
15. East R, Kain W. Prediction of grass grub, *Costelytra zealandica*,(Coleoptera: Scarabaeidae) populations. New Zealand Entomologist 1982; 7: 222-226.
16. Chapman RB, Jackson T. A review of insecticide use on pastures and forage crops in New Zealand. http://agpestconz/ 2010.
17. Jackson TA. A novel bacterium for control of grass grub. In: Biological control: A global perspective. Wallingford, U.K.: CABI publishing; 2007. p. 160-168.
18. Grimont PAD, Jackson TA, Ageron E, Noonan MJ. *Serratia entomophila* sp. nov. associated with amber disease in the New Zealand grass grub *Costelytra zealandica*. International Journal of Systematic Bacteriology 1988; 38: 1-6.
19. Ferguson CM, Barratt BIP, Bell N, Goldson SL, Hardwick S, Jackson M, et al. Quantifying the economic cost of invertebrate pests to New Zealand's pastoral industry. New Zealand Journal of Agricultural Research 2018: 1-61.
20. Kelsey JM. Grass-grub and grass caterpillar control. New Zealand Journal of Agricultural Research 1951: 113-122.

21. Perrott DCF. Susceptibility and tolerance to DDT and other insecticides in larvae of *Costelytra zealandica* (white) (Melolonthinae, Scarabaeidae). New Zealand Journal of Agricultural Research 1966; 9: 68-77.
22. Henzell RF, Skinner RA, Clements RO. Insecticides for control of adult grass grub, *Costelytra zealandica* (White). New Zealand Journal of Agricultural Research 1983; 26: 129-133.
23. Lauren DR, Henzell RF, Wrenn NR. Grass grub (*Costelytra zealandica*) population trends following insecticide applications. New Zealand Journal of Agricultural Research 1990; 33: 159-163.
24. Wu G, Wang Y, Wang JN, Chen XZ, Hu QX, Yang YF, et al. Vitality and stability of insecticide resistance in adult *Propylaea japonica* (Coleoptera: Coccinellidae). Journal of Insect Science 2018; 18.
25. Zydenbos SM, Townsend RJ, Lane PMS, Mansfield S, O'Callaghan M, Koten Cv, et al. Effect of *Serratia entomophila* and diazinon applied with seed against grass grub populations on the North Island volcanic plateau. New Zealand Plant Protection 2016; 69: 86-93.
26. Akay-Demir AE, Dilek FB, Yetis U. A new screening index for pesticides leachability to groundwater. Journal of Environmental Management 2019; 231: 1193-1202.
27. Iwafune T. Studies on the behavior and ecotoxicity of pesticides and their transformation products in a river. Journal of Pesticide Science 2018; 43: 297-304.
28. Kwon YK, Wee SH, Kim JH. Pesticide poisoning events in wild birds in Korea from 1998 to 2002. Journal of Wildlife Diseases 2004; 40: 737-740.
29. Muñoz-Arnanz J, Chirife AD, Galletti-Vernazzani B, Cabrera E, Sironi M, Millán J, et al. First assessment of persistent organic pollutant contamination in blubber of Chilean blue whales from Isla de Chiloé, southern Chile. Science of The Total Environment 2019; 650: 1521-1528.
30. Pisa LW, Amaral-Rogers V, Belzunces LP, Bonmatin JM, Downs CA, Goulson D, et al. Effects of neonicotinoids and fipronil on non-target invertebrates. Environmental Science and Pollution Research 2015; 22: 68-102.
31. Zaller JG, König N, Tiefenbacher A, Muraoka Y, Querner P, Ratzenböck A, et al. Pesticide seed dressings can affect the activity of various soil organisms and reduce decomposition of plant material. BMC Ecology 2016; 16: 37-37.
32. Sak ZHA, Kurtuluş Ş, Ocakli B, TÖReyİN ZN, Bayhan İ, YeŞİLnacar Mİ, et al. Respiratory symptoms and pulmonary functions before and after pesticide application in cotton farming. Annals of Agricultural and Environmental Medicine 2018; 25: 701-707.
33. Mrema EJ, Ngowi AV, Kishinhi SS, Mamuya SH. Pesticide exposure and health problems among female horticulture workers in Tanzania. Environmental Health Insights 2017; 11: 11-13.
34. Ma Y, He X, Qi K, Wang T, Qi Y, Cui L, et al. Effects of environmental contaminants on fertility and reproductive health. Journal of Environmental Sciences 2019; 77: 210-217.
35. Aktar MW, Sengupta D, Chowdhury A. Impact of pesticides use in agriculture: their benefits and hazards. Interdisciplinary Toxicology 2009; 2: 1-12.
36. Flaskos J, Harris W, Sachana M, Munoz D, Tack J, Hargreaves AJ. The effects of diazinon and cypermethrin on the differentiation of neuronal and glial cell lines. Toxicology and Applied Pharmacology 2007; 219: 172-180.
37. Lacey LA, Grzywacz D, Shapiro-Ilan DI, Frutos R, Brownbridge M, Goettel MS. Insect pathogens as biological control agents: Back to the future. Journal of Invertebrate Pathology 2015; 132: 1-41.
38. Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, et al. Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. Cell 2016; 166: 1397-1410.e1316.
39. Seiler H, Busse M. The yeasts of cheese brines. International Journal of Food Microbiology 1990; 11: 289-303.
40. Dave RI, Shah NP. Viability of yoghurt and probiotic bacteria in yoghurts made from commercial starter cultures. International Dairy Journal 1997; 7: 31-41.
41. Jiang Q, Ghim D, Cao S, Tadepalli S, Liu K, Kwon H, et al. Photothermally active reduced graphene oxide/bacterial nanocellulose composites as biofouling-resistant ultrafiltration membranes. environmental science & technology 2019; 53: 412-421.
42. Khan MI, Shin JH, Kim JD. The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. Microbial Cell Factories 2018; 17: 36-36.

43. Abishek MP, Patel J, Rajan AP. Algae oil: a sustainable renewable fuel of future. Biotechnology Research International 2014; 2014: 272814-272814.
44. Glare TR, Gwynn RL, Moran-Diez ME. Development of biopesticides and future opportunities. In: Microbial-based biopesticides: Methods and protocols. New York, NY: Springer New York; 2016. p. 211-221.
45. Ehlers RU. Current and future use of nematodes in biocontrol: Practice and commercial aspects with regard to regulatory policy issues. Biocontrol Science and Technology 1996; 6: 303-316.
46. Dolinski C, Choo HY, Duncan LW. Grower acceptance of entomopathogenic nematodes: case studies on three continents. Journal of Nematology 2012; 44: 226-235.
47. ffrench-Constant RH, Bowen DJ. Novel insecticidal toxins from nematode-symbiotic bacteria. Cellular and Molecular Life Sciences 2000; 57: 828-833.
48. Gatsogiannis C, Lang AE, Meusch D, Pfaumann V, Hofnagel O, Benz R, et al. A syringe-like injection mechanism in Photorhabdus luminescens toxins. Nature 2013; 495: 520-523.
49. Dillman AR, Sternberg PW. Entomopathogenic nematodes. Current biology 2012; 22: R430-R431.
50. Vega FE, Posada F, Catherine-Aime M, Pava-Ripoll M, Infante F, Rehner SA. Entomopathogenic fungal endophytes. Biological Control 2008; 46: 72-82.
51. Vega FE, Goettel MS, Blackwell M, Chandler D, Jackson MA, Keller S, et al. Fungal entomopathogens: new insights on their ecology. Fungal Ecology 2009; 2: 149-159.
52. Khare E, Mishra J, Arora NK. Multifaceted Interactions Between endophytes and plant: developments and prospects. Frontiers in Microbiology 2018; 9: 2732-2732.
53. Bamisile BS, Dash CK, Akutse KS, Keppanan R, Wang L. Fungal endophytes: Beyond herbivore management. Frontiers in Microbiology 2018; 9.
54. Pappas ML, Liapoura M, Papantoniou D, Avramidou M, Kavroulakis N, Weinhold A, et al. The beneficial endophytic fungus *Fusariumsolani* strain K alters tomato responses against spider mites to the benefit of the plant. Frontiers in Plant Science 2018; 9.
55. Jurat-Fuentes JL, Jackson TA. Chapter 8 - Bacterial Entomopathogens. In: Insect Pathology. San Diego: Academic Press; 2012. p. 265-349.
56. Berliner E. Uber die Schlaffsucht der Mehlmottenraupe. Z ges Getreidew 1911; 3: 63-70.
57. Berliner E. Über die Schlaffsucht der Mehlmottenraupe (Ephestia kühniella Zell.) und ihren Erreger Bacillus thuringiensis n. sp. Zeitschrift für angewandte Entomologie 1915; 2: 29-56.
58. Palma L, Muñoz D, Berry C, Murillo J, Caballero P. *Bacillus thuringiensis* toxins: an overview of their biocidal activity. Toxins 2014; 6: 3296-3325.
59. Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, et al. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiology and Molecular Biology Reviews 1998; 62: 775-806.
60. Goldberg LJ, Margalit J. A bacterial spore demonstrating rapid larvicidal activity against *Anopheles sergentii*, *Uranotaenia unguiculata*, *Culex univitattus*, *Aedes aegypti* and *Culex pipiens*. Mosquito News 1977; 37: 355-358.
61. Méric G, Mageiros L, Pascoe B, Woodcock DJ, Mourkas E, Lamble S, et al. Lineage-specific plasmid acquisition and the evolution of specialized pathogens in *Bacillus thuringiensis* and the Bacillus cereus group. Molecular Ecology 2018; 27: 1524-1540.
62. González JM, Jr., Brown BJ, Carlton BC. Transfer of *Bacillus thuringiensis* plasmids coding for delta-endotoxin among strains of *B. thuringiensis* and *B. cereus*. Proceedings of the National Academy of Sciences of the United States of America 1982; 79: 6951-6955.
63. Méric G, Mageiros L, Pascoe B, Woodcock DJ, Mourkas E, Lamble S, et al. Lineage-specific plasmid acquisition and the evolution of specialized pathogens in *Bacillus thuringiensis* and the Bacillus cereus group. Molecular Ecology 2018; 27: 1524-1540.
64. Romeis J, Meissle M, Bigler F. Transgenic crops expressing *Bacillus thuringiensis* toxins and biological control. Nature Biotechnology 2006; 24: 63.
65. Jackson TA, Pearson JF, O'Callaghan M, Mahanty HK, Willocks M. Pathogen to product - development of *Serratia entomophila* (Enterobacteriacae) as a commercial biological control agent for the New Zealand grass grub (*Costelytra zealandica*). In: Pathogens in Scarab Pest Management. Andover, U.K.: Intercept; 1992.

66. Adeolu M, Alnajar S, Naushad S, R SG. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. International Journal of Systematic and Evolutionary Microbiology 2016; 66: 5575-5599.
67. Glare TR, O'Callaghan M, Wigley PJ. Checklist of naturally occurring entomopathogenic microbes and nematodes in New Zealand. New Zealand Journal of Zoology 1993; 20: 95-120.
68. Jackson T. Biological control of grass grub in Canterbury. Proceedings of the New Zealand Grassland Association 1990; 52: 217-220.
69. Hurst MRH, Beattie A, Jones SA, Laugraud A, van Koten C, Harper L. Characterization of *Serratia proteamaculans* strain AGR96X encoding an anti-feeding prophage (tailocin) with activity against grass grub (*Costelytra giveni*) and manuka beetle (*Pyronota* spp.) larvae. Applied and Environmental Microbiology 2018.
70. Trought TET, Jackson TA, French RA. Incidence and transmission of a disease of grass grub (*Costelytra zealandica*) in Canterbury. New Zealand Journal of Experimental Agriculture 1982; 10: 79-82.
71. Grimes M, Hennerty AJ, editors. A study of bacteria belonging to the sub-genus *Aerobacter* 1931: Royal Dublin Society.
72. Grimont PAD, Grimont F, Irino K. Biochemical characterization of *Serratia liquefaciens* sensu stricto, *Serratia proteamaculans*, and *Serratia grimesii* sp. nov. Current Microbiology 1982; 7: 69-74.
73. Bascomb S, Lapage SP, Willcox WR, Curtis MA. Numerical classification of the tribe *Klebsielleae*. The Journal of General Microbiology 1971; 66: 279-295.
74. Grimont PA, Grimont F, Starr MP. *Serratia proteamaculans* (Paine and Stansfield) comb. nov., a Senior Subjective Synonym of *Serratia liquefaciens* (Grimes and Hennerty) Bascomb et al. International Journal of Systematic and Evolutionary Microbiology 1978; 28: 503-510.
75. Grimont PA, Grimont F, De Rosnay HL. Taxonomy of the genus Serratia. The Journal of General Microbiology 1977; 98: 39-66.
76. Grimont F, Grimont PAD. The genus *serratia*. in: The Prokaryotes: Volume 6: Proteobacteria: Gamma Subclass 2006: 219-244.
77. O'Callaghan M, Jackson TA, Mahanty HK. Selection, development and testing of phage-resistant strains of *Serratia entomophila* for grass grub control. Biocontrol Science and Technology 1992; 2: 297-305.
78. Pearson JF, Jackson TA. Quality control management of the grass grub microbial control product, Invade®. Proceedings Agronomy Society of New Zealand 1995; 25.
79. Jackson TA, Glare TR. Use of pathogens in scarab pest management. Andover, Hampshire: Intercept; 1992. 298 p. p.
80. Jackson T, Glare T, O'Callaghan M. Pathotypic boundaries for *Serratia* ssp. causing amber disease in the New Zealand grass grub, *Costelytra zealandica*. Proceedings of the 3rd European Meeting of Microbial Control of Pests 1991: p. 148-152.
81. Jackson TA, Huger AM, Glare TR. Pathology of amber disease in the New Zealand grass grub *Costelytra zealandica* (Coleoptera: Scarabaeidae). Journal of Invertebrate Pathology 1993; 61: 123-130.
82. Zheng J, Peng D, Ruan L, Sun M. Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the *Bacillus cereus* group. BMC Evolutionary Biology 2013; 13: 262-262.
83. Hurst MRH, Becher SA, O'Callaghan M. Nucleotide sequence of the *Serratia entomophila* plasmid pADAP and the *Serratia proteamaculans* pU143 plasmid virulence associated region. Plasmid 2011; 65: 32-41.
84. Claus H, Jackson TA, Filip Z. Characterization of *Serratia entomophila* strains by genomic DNA fingerprints and plasmid profiles. Microbiological Research 1995; 150: 159-166.
85. Grkovic S, Glare TR, Jackson TA, Corbett GE. Genes essential for amber disease in grass grubs are located on the large plasmid found in *Serratia entomophila* and *Serratia proteamaculans*. Applied and Environmental Microbiology 1995; 61: 2218-2223.

86. Glare TR, Hurst MRH, Grkovic S. Plasmid transfer among several members of the family Enterobacteriaceae increases the number of species capable of causing experimental amber disease in grass grub. FEMS Microbiology Letters 1996; 139: 117-120.
87. Johnson VW, Pearson J, Jackson TA. Formulation of *Serratia entomophila* for biological control of grass grub. New Zealand Plant Protection 2001; 54: 125-127.
88. Townsend RJ, Ferguson CM, Proffitt JR, Slay MWA, Swaminathan J, Day S, Gerard E, et al. Establisment of *Serratia entomophila* after application of a new formulation for grass grub control. New Zealand Plant Protection 2004; 57: 310-313.
89. Trought TET, Jackson TA, French RA. Incidence and transmission of a disease of grass grub (*Costelytra zealandica*) in Canterbury. New Zealand Journal of Experimental Agriculture 1982; 10: 79-82.
90. Dodd SJ, Hurst MRH, Glare TR, O'Callaghan M, Ronson CW. Occurrence of *sep* Insecticidal Toxin Complex Genes in *Serratia* spp. and *Yersinia frederiksenii*. Applied and Environmental Microbiology 2006; 72: 6584-6592.
91. Fabricius JC. Systema entomologiae: sistens insectorvm classes, ordines, genera, species, adiectis synonymis, locis, descriptionibvs, observationibvs / Io. Christ. Fabricii. Flensbvrgi et Lipsiae: In Officina Libraria Kortii; 1775.
92. Stewart KM. Preliminary observations on the biology of a manuka chafer, *Pyronota* sp. (Coleoptera: Scarabaeidae) in Otago. New Zealand Entomologist 1987; 9: 60-63.
93. Dunbar JE, Hateley PW, Townsend RJ, Zydenbos SM, Jackson TA. Assessing damage by manuka beetles (*Pyronota* spp.) in “flipped” West Coast pastures. New Zealand Plant Protection 2012; 65.
94. Dodd SJ. Horizontal transfer of plasmidborne insecticidal toxin genes of *Serratia* species (doctoral thesis). Dunedin 2003.
95. Clarke M, Maddera L, Harris RL, Silverman PM. F-pili dynamics by live-cell imaging. Proceedings of the National Academy of Sciences of the United States of America 2008; 105: 17978-17981.
96. Pelicic V. Type IV pili: e pluribus unum? Molecular Microbiology 2008; 68: 827-837.
97. Wong JJW, Lu J, Glover JNM. Relaxosome function and conjugation regulation in F-like plasmids – a structural biology perspective. Molecular Microbiology 2012; 85: 602-617.
98. Del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R. Replication and control of circular bacterial plasmids. Microbiology and molecular biology reviews 1998; 62: 434-464.
99. Novick RP. Plasmid incompatibility. Microbiological Reviews 1987; 51: 381-395.
100. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing. Journal of Microbiological Methods 2005; 63: 219-228.
101. Ruiz-Masó JA, MachóN C, Bordanaba-Ruiseco L, Espinosa M, Coll M, Del Solar G. Plasmid rolling-circle replication. Microbiology Spectrum 2015; 3.
102. Khan SA. Rolling-circle replication of bacterial plasmids. Microbiology and molecular biology reviews 1997; 61: 442-455.
103. Llosa M, Gomis-Rüth FX, Coll M, Cruz Fdl. Bacterial conjugation: a two-step mechanism for DNA transport. Molecular Microbiology 2002; 45: 1-8.
104. Jheeta S. Horizontal gene transfer and its part in the reorganisation of genetics during the LUCA Epoch. Life 2013; 3: 518-523.
105. Penades JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. Bacteriophage-mediated spread of bacterial virulence genes. Current Opinion in Microbiology 2015; 23: 171-178.
106. Domingues S, Nielsen KM. Membrane vesicles and horizontal gene transfer in prokaryotes. Current Opinion in Microbiology 2017; 38: 16-21.
107. Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. Nature Reviews Microbiology 2012; 10: 472-482.
108. Lederberg J. The transformation of genetics by DNA: an anniversary celebration of Avery, MacLeod and McCarty (1944). Genetics 1994; 136: 423-426.
109. Norman A, Hansen LH, Sorensen SJ. Conjugative plasmids: vessels of the communal gene pool. Philosophical Transactions of the Royal Society B: Biological Sciences 2009; 364: 2275-2289.
110. Silver S. Bacterial resistances to toxic metal ions--a review. Gene 1996; 179: 9-19.

111. Schroeder GN, Hilbi H. Molecular Pathogenesis of *Shigella* spp.: Controlling host cell signaling, invasion, and death by Type III secretion. Clinical Microbiology Reviews 2008; 21: 134-156.
112. Naik OA, Shashidhar R, Rath D, Bandekar JR, Rath A. Characterization of multiple antibiotic resistance of culturable microorganisms and metagenomic analysis of total microbial diversity of marine fish sold in retail shops in Mumbai, India. Environmental Science and Pollution Research 2017; 25: 6228-6239.
113. Dolejska M, Papagiannitsis CC. Plasmid-mediated resistance is going wild. Plasmid 2018.
114. Rangasamy K, Athiappan M, Devarajan N, Samykannu G, Parray JA, Aruljothi KN, et al. Pesticide degrading natural multidrug resistance bacterial flora. Microbial Pathogenesis 2017; 114: 304-310.
115. Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Muller U, Heijne W, et al. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biology and Evolution 2010; 2: 212-224.
116. Kinashi H. Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. The Journal of Antibiotics 2011; 64: 19-25.
117. Lobkovsky AE, Wolf YI, Koonin EV. Estimation of prokaryotic supergenome size and composition from gene frequency distributions. BMC Genomics 2014; 15: S14.
118. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, et al. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 2006; 314: 267.
119. Harrison PW, Lower RP, Kim NK, Young JP. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends in Microbiology 2010; 18: 141-148.
120. Landeta C, Davalos A, Cevallos MA, Geiger O, Brom S, Romero D. Plasmids with a chromosome-like role in *rhizobia*. Journal of Bacteriology 2011; 193: 1317-1326.
121. San Millan A, MacLean RC. Fitness costs of plasmids: a limit to plasmid transmission. Microbiology Spectrum 2017; 5.
122. Baltrus DA. Exploring the costs of horizontal gene transfer. Trends in Ecology & Evolution 2013; 28: 489-495.
123. Vogwill T, MacLean RC. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. Evolutionary Applications 2015; 8: 284-295.
124. Rizvi SMA, Prajapati HK, Ghosh SK. The 2 micron plasmid: a selfish genetic element with an optimized survival strategy within *Saccharomyces cerevisiae*. Current Genetics 2018; 64: 25-42.
125. Werren JH. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proceedings of the National Academy of Sciences of the United States of America 2011; 108: 10863-10870.
126. Shintani M, Takahashi Y, Tokumaru H, Kadota K, Hara H, Miyakoshi M, et al. Response of the *Pseudomonas* host chromosomal transcriptome to carriage of the IncP-7 plasmid pCAR1. Environmental Microbiology 2010; 12: 1413-1426.
127. Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. Current Biology 2015; 25: 2034-2039.
128. Cheng X, de Bruijn I, van der Voort M, Loper JE, Raaijmakers JM. The Gac regulon of *Pseudomonas fluorescens* SBW25. Environmental Microbiology Reports 2013; 5: 608-619.
129. San Millan A, Toll-Riera M, Qi Q, Betts A, Hopkinson RJ, McCullagh J, et al. Integrative analysis of fitness and metabolic effects of plasmids in *Pseudomonas aeruginosa* PAO1. The ISME Journal 2018; 12: 3014-3024.
130. Hall JPJ, Harrison E, Lilley AK, Paterson S, Spiers AJ, Brockhurst MA. Environmentally co-occurring mercury resistance plasmids are genetically and phenotypically diverse and confer variable context-dependent fitness effects. Environmental Microbiology 2015; 17: 5008-5022.
131. Attéré SA, Vincent AT, Paccaud M, Frenette M, Charette SJ. The role for the small cryptic plasmids as moldable vectors for genetic innovation in *Aeromonas salmonicida* subsp. salmonicida. Frontiers in Genetics 2017; 8.
132. Sorensen SJ, Bailey M, Hansen LH, Kroer N, Wuertz S. Studying plasmid horizontal transfer in situ: a critical review. Nature Reviews Microbiology 2005; 3: 700-710.
133. Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. Why genes evolve faster on secondary chromosomes in bacteria. PLOS Computational Biology 2010; 6: e1000732.

134. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. Proceedings of the National Academy of Sciences of the United States of America 2005; 102: 14338.
135. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 2008; 134: 341-352.
136. Pál C, Papp B, Lercher MJ. An integrated view of protein evolution. Nature Reviews Genetics 2006; 7: 337-348.
137. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, et al. Functional genomic analysis of the rates of protein evolution. Proceedings of the National Academy of Sciences of the United States of America 2005; 102: 5483.
138. Rodriguez-Beltran J, Hernandez-Beltran JCR, DelaFuente J, Escudero JA, Fuentes-Hernandez A, MacLean RC, et al. Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation. Nature ecology & evolution 2018; 2: 873-881.
139. Melnyk RA, Haney CH. Plasmid-powered evolutionary transitions. eLife 2017; 6.
140. Lili LN, Britton NF, Feil EJ. The persistence of parasitic plasmids. Genetics 2007; 177: 399-405.
141. Iranzo J, Puigbo P, Lobkovsky AE, Wolf YI, Koonin EV. Inevitability of genetic parasites. Genome Biology and Evolution 2016; 8: 2856-2869.
142. Yang QE, Walsh TR. Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. FEMS microbiology reviews 2017; 41: 343-353.
143. Diaz-Orejas R, Espinosa M, Yeo CC. The Importance of the expendable: toxin-antitoxin gGenes in plasmids and chromosomes. Frontiers in Microbiology 2017; 8: 1479.
144. Gerdes K, Rasmussen PB, Molin S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proceedings of the National Academy of Sciences 1986; 83: 3116.
145. Majeed H, Lampert A, Ghazaryan L, Gillor O. The weak shall inherit: bacteriocin-mediated interactions in bacterial populations. PLoS One 2013; 8: e63837.
146. Inglis RF, Bayramoglu B, Gillor O, Ackermann M. The role of bacteriocins as selfish genetic elements. Biology Letters 2013; 9: 20121173.
147. Page R, Peti W. Toxin-antitoxin systems in bacterial growth arrest and persistence. Nature Chemical Biology 2016; 12: 208-214.
148. Wood TK, Knabel SJ, Kwan BW. Bacterial persister cell formation and dormancy. Applied and Environmental Microbiology 2013; 79: 7116-7121.
149. Kothari A, Wu Y-W, Chandonia J-M, Charrier M, Rajeev L, Rocha AM, et al. Large circular plasmids from groundwater plasmidomes span multiple incompatibility groups and are enriched in multimetal resistance genes. mBio 2019; 10: e02899-02818.
150. Ferreira C, Bogas D, Bikarolla SK, Varela AR, Frykholm K, Linheiro R, et al. Genetic variation in the conjugative plasmidome of a hospital effluent multidrug resistant *Escherichia coli* strain. Chemosphere 2019; 220: 748-759.
151. Szczepanowski R, Bekel T, Goesmann A, Krause L, Krömeke H, Kaiser O, et al. Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. Journal of Biotechnology 2008; 136: 54-64.
152. Bleicher A, Schöfl G, Rodicio MdR, Saluz HP. The plasmidome of a *Salmonella enterica* serovar Derby isolated from pork meat. Plasmid 2013; 69: 202-210.
153. Brolund A, Franzén O, Melefors O, Tegmark-Wisell K, Sandegren L. Plasmidome-analysis of ESBL-producing *escherichia coli* using conventional typing and high-throughput sequencing. PLoS One 2013; 8: e65793-e65793.
154. Carroll AC, Wong A. Plasmid persistence: costs, benefits, and the plasmid paradox. Canadian Journal of Microbiology 2018; 64: 293-304.
155. Jain R, Rivera MC, Moore JE, Lake JA. Horizontal gene transfer accelerates genome innovation and evolution. Molecular Biology and Evolution 2003; 20: 1598-1602.
156. Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends in Microbiology 2012; 20: 262-267.

157. Wang X, Liu D, Luo Y, Zhao L, Liu Z, Chou M, et al. Comparative analysis of *rhizobial* chromosomes and plasmids to estimate their evolutionary relationships. Plasmid 2018; 96-97: 13-24.
158. Seth-Smith HMB, Harris SR, Persson K, Marsh P, Barron A, Bignell A, et al. Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. BMC Genomics 2009; 10: 239.
159. Brantl S. Antisense-RNA mediated control of plasmid replication – pIP501 revisited. Plasmid 2015; 78: 4-16.
160. Edwards RA, Schifferli DM, Maloy SR. A role for *Salmonella fimbriae* in intraperitoneal infections. Proceedings of the National Academy of Sciences of the United States of America 2000; 97: 1258-1262.
161. Hurst MRH, Beard SS, Jackson TA, Jones SM. Isolation and characterization of the *Serratia entomophila* antifeeding prophage. FEMS Microbiology Letters 2007; 270: 42-48.
162. Dobrindt U, Blum-Oehler G, Nagy G, Schneider G, Johann A, Gottschalk G, et al. Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. Infection and Immunity 2002; 70: 6365-6372.
163. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clinical Microbiology Reviews 2004; 17: 14-56.
164. Peed L, Parker AC, Smith CJ. Genetic and functional analyses of the mob operon on conjugative transposon CTn341 from *Bacteroides* spp. Journal of Bacteriology 2010; 192: 4643-4650.
165. Henderson D, Meyer RJ. The primase of broad-host-range plasmid R1162 is active in conjugal transfer. Journal of Bacteriology 1996; 178: 6888.
166. Perwez T, Meyer R. MobB protein stimulates nicking at the R1162 origin of transfer by increasing the proportion of complexed plasmid DNA. Journal of Bacteriology 1996; 178: 5762-5767.
167. Kurenbach B, Grothe D, Farías ME, Szewzyk U, Grohmann E. The tra region of the conjugative plasmid pIP501 Is organized in an operon with the first gene encoding the relaxase. Journal of Bacteriology 2002; 184: 1801.
168. Zechner EL, Lang S, Schildbach JF. Assembly and mechanisms of bacterial type IV secretion machines. Philosophical Transactions of the Royal Society B: Biological Sciences 2012; 367: 1073-1087.
169. Grohmann E, Muth G, Espinosa M. Conjugative Plasmid transfer in gram-positive bacteria. Microbiology and Molecular Biology Reviews 2003; 67: 277.
170. Miki T, Horiuchi T, Willetts NS. Identification and characterization of four new tra cistrons on the *E. coli* K12 sex factor F. Plasmid 1978; 1: 316-323.
171. Harris RL, Silverman PM. Tra Proteins Characteristic of F-Like Type IV secretion systems constitute an interaction group by yeast two-hybrid analysis. Journal of Bacteriology 2004; 186: 5480-5485.
172. Lawley TD, Gordon GS, Wright A, Taylor DE. Bacterial conjugative transfer: visualization of successful mating pairs and plasmid establishment in live *Escherichia coli*. Molecular Microbiology 2002; 44: 947-956.
173. Hu B, Khara P, Christie PJ. Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*. Proceedings of the National Academy of Sciences 2019; 116: 14222-14227.
174. Costa TRD, Ilangovan A, Ukleja M, Redzej A, Santini JM, Smith TK, et al. Structure of the bacterial sex F pilus reveals an assembly of a stoichiometric protein-phospholipid complex. Cell 2016; 166: 1436-1444.e1410.
175. Panicker MM, Minkley EG, Jr. DNA transfer occurs during a cell surface contact stage of F sex factor-mediated bacterial conjugation. Journal of Bacteriology 1985; 162: 584-590.
176. Touhami A, Jericho MH, Boyd JM, Beveridge TJ. Nanoscale characterization and determination of adhesion forces of *Pseudomonas aeruginosa* pili by using atomic force microscopy. Journal of Bacteriology 2006; 188: 370-377.
177. Craig L, Forest KT, Maier B. Type IV pili: dynamics, biophysics and functional consequences. Nature Reviews Microbiology 2019; 17: 429-440.
178. Piepenbrink KH. DNA Uptake by Type IV filaments. Frontiers in Molecular Biosciences 2019; 6.

179. Denise R, Abby SS, Rocha EPC. Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. PLOS Biology 2019; 17: e3000390.
180. Babic A, Lindner AB, Vulic M, Stewart EJ, Radman M. Direct visualization of horizontal gene transfer. Science 2008; 319: 1533-1536.
181. Waksman G. From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. EMBO reports 2019; 20: e47012.
182. Aviv G, Tsyba K, Steck N, Salmon-Divon M, Cornelius A, Rahav G, et al. A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *Salmonella enterica* serovar Infantis strain. Environmental Microbiology 2014; 16: 977-994.
183. Knight SD, Bouckaert J. Structure, function, and assembly of Type 1 fimbriae. In: Glycoscience and microbial adhesion. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 67-107.
184. Müller CM, Åberg A, Straseviçiene J, Emődy L, Uhlin BE, Balsalobre C. Type 1 fimbriae, a colonization factor of uropathogenic *escherichia coli*, are controlled by the metabolic sensor CRP-cAMP. PLoS Pathogens 2009; 5: e1000303.
185. Connell I, Agace W, Klemm P, Schembri M, Mărild S, Svanborg C. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. Proceedings of the National Academy of Sciences of the United States of America 1996; 93: 9827-9832.
186. Althouse C, Patterson S, Fedorka-Cray P, Isaacson RE. Type 1 fimbriae of *Salmonella enterica* serovar Typhimurium bind to enterocytes and contribute to colonization of swine in vivo. Infection and immunity 2003; 71: 6446-6452.
187. Avalos Vizcarra I, Hosseini V, Kollmannsberger P, Meier S, Weber SS, Arnoldini M, et al. How type 1 fimbriae help *Escherichia coli* to evade extracellular antibiotics. Scientific Reports 2016; 6: 18109.
188. Chandra H, Khandelwal P, Khattri A, Banerjee N. Type 1 fimbriae of insecticidal bacterium *Xenorhabdus nematophila* is necessary for growth and colonization of its symbiotic host nematode *Steinernema carpocapsiae*. Environmental Microbiology 2008; 10: 1285-1295.
189. ffrench-Constant R, Waterfield N. An ABC Guide to the Bacterial Toxin Complexes. In: Advances in Applied Microbiology. 58: Academic Press; 2005. p. 169-183.
190. Hurst MRH, Glare TR, Jackson TA, Ronson CW. Plasmid-Located Pathogenicity Determinants of *Serratia entomophila*, the Causal Agent of Amber Disease of Grass Grub, Show Similarity to the Insecticidal Toxins of *Photorhabdus luminescens*. Journal of Bacteriology 2000; 182: 5127-5138.
191. Yang G, Dowling AJ, Gerike U, ffrench-Constant RH, Waterfield NR. Photorhabdus virulence cassettes confer injectable insecticidal activity against the wax moth. Journal of Bacteriology 2006; 188: 2254-2261.
192. Hurst MRH, O'Callaghan M, Glare TR. Peripheral sequences of the *Serratia entomophila* pADAP virulence-associated region. Plasmid 2003; 50: 213-229.
193. Waterfield N, Hares M, Yang G, Dowling A, ffrench-Constant R. Potentiation and cellular phenotypes of the insecticidal Toxin complexes of *Photorhabdus* bacteria. Cell Microbiol 2005; 7: 373-382.
194. Heymann JB, Bartho JD, Rybakova D, Venugopal HP, Winkler DC, Sen A, et al. Three-dimensional structure of the toxin-delivery particle antifeeding prophage of *Serratia entomophila*. Journal of Biological Chemistry 2013; 288: 25276-25284.
195. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biology Direct 2012; 7: 18.
196. Croll D, Lendenmann MH, Stewart E, McDonald BA. The Impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 2015; 201: 1213-1228.
197. Busby JN, Panjikar S, Landsberg MJ, Hurst MRH, Lott JS. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. Nature 2013; 501: 547-550.
198. Hurst MRH, Glare TR, Jackson TA. Cloning *Serratia entomophila* antifeeding genes—a putative defective prophage active against the grass grub *Costelytra zealandica*. Journal of Bacteriology 2004; 186: 5116-5128.

199. Jackson TA. Amber disease reduces trypsin activity in midgut of *Costelytra zealandica* (Coleoptera; Scarabaeidae) larvae. Journal of Invertebrate Pathology 1995; 65: 68-69.
200. Gatehouse HS, Tan B, Christeller JT, Hurst MRH, Marshall SD, Jackson TA. Phenotypic changes and the fate of digestive enzymes during induction of amber disease in larvae of the New Zealand grass grub (*Costelytra zealandica*). Journal of Invertebrate Pathology 2009; 101: 215-221.
201. Rybakova D, Radjainia M, Turner A, Sen A, Mitra AK, Hurst MRH. Role of antifeeding prophage (Afp) protein Afp16 in terminating the length of the Afp tailocin and stabilizing its sheath. Molecular Microbiology 2013; 89: 702-714.
202. Desfosses A, Venugopal HP, Joshi T, Felix J, Jessop M, Jeong H, et al. Atomic structures of an entire contractile injection system in both the extended and contracted states. Nature Microbiology 2019.
203. Rybakova D, Schramm P, Mitra AK, Hurst MRH. Afp14 is involved in regulating the length of anti-feeding prophage (Afp). Molecular Microbiology 2015; 96: 815-826.
204. Ghequire MGK, De Mot R. The tailocin tale: peeling off phage tails. Trends in Microbiology 2015; 23: 587-590.
205. Sen A, Rybakova D, Hurst MRH, Mitra AK. Structural study of the *Serratia entomophila* antifeeding prophage: three-dimensional structure of the helical sheath. Journal of Bacteriology 2010; 192: 4522-4525.
206. Coulthurst SJ. The Type VI secretion system - a widespread and versatile cell targeting system. Research in Microbiology 2013; 164: 640-654.
207. Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S, et al. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Molecular Microbiology 2000; 38: 213-231.
208. Tan B, Jackson TA, Hurst MRH. Virulence of *Serratia* strains against *Costelytra zealandica*. Applied and Environmental Microbiology 2006; 72: 6417-6418.
209. Luria SE, Burrous JW. Hybridization between *Escherichica coli* and *Shigella*. Journal of Bacteriology 1957; 74: 461-476.
210. Chang AC, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. Journal of Bacteriology 1978; 134: 1141-1156.
211. Prentki P, Krisch HM. In vitro insertional mutagenesis with a selectable DNA fragment. Gene 1984; 29: 303-313.
212. Barbier M, Damron FH. Rainbow vectors for broad-range bacterial fluorescence labeling. PLoS One 2016; 11: e0146827.
213. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 1996; 173: 33-38.
214. Riedel T, Rohlfs M, Buchholz I, Wagner-Döbler I, Reck M. Complete sequence of the suicide vector pJP5603. Plasmid 2013; 69: 104-107.
215. Durfee T, Nelson R, Baldwin S, Plunkett 3rd G, Burland V, Mau B, et al. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. Journal of Bacteriology 2008; 190: 2597-2606.
216. Thoma S, Schobert M. An improved *Escherichia coli* donor strain for diparental mating. FEMS Microbiology Letters 2009; 294: 127-132.
217. Blomfield IC, McClain MS, Princ JA, Calie PJ, Eisenstein BI. Type 1 fimbriation and fimE mutants of *Escherichia coli* K-12. Journal of bacteriology 1991; 173: 5298-5307.
218. Glare TR, Corbett GE, Sadler TJ. Association of a large plasmid with amber disease of the New Zealand grass grub, *Costelytra zealandica*, caused by *Serratia entomophila* and *Serratia proteamaculans*. Journal of Invertebrate Pathology 1993; 62: 165-170.
219. Kado CI, Liu ST. Rapid procedure for detection and isolation of large and small plasmids. Journal of Bacteriology 1981; 145: 1365-1373.
220. Atanassov II, Atanassov II, Etchells JP, Turner SR. A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. Plant Methods 2009; 5: 14-14.
221. Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, Taheri-Talesh N, et al. Fusion PCR and gene targeting in *Aspergillus nidulans*. Nature Protocols 2006; 1: 3111-3120.

222. FlowJo T. FlowJo. v10.6.1 ed. Oregon, United States of America: FlowJo, LLC; 2019.
223. SPECTROstar Nano MARS. v2.10 R5 ed. Ortenberg, Germany: BMG LABTECH; 2018.
224. Hockett KL, Baltrus DA. Use of the soft-agar overlay technique to screen for bacterially produced inhibitory compounds. JoVE 2017: e55064.
225. Chevallet M, Luche S, Rabilloud T. Silver staining of proteins in polyacrylamide gels. Nature Protocols 2006; 1: 1852-1858.
226. J.F S, W. Russell D. Molecular cloning: A laboratory manual 2001.
227. Thain D, Tannenbaum T, Livny M. Distributed computing in practice: the Condor experience. Concurrency and Computation: Practice and Experience 2005; 17: 323-356.
228. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research 2010; 38: 1767-1771.
229. Trim Galore. v0.6.2 ed. Cambridge, England: Babraham Bioinformatics; 2019.
230. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research 1998; 8: 175-185.
231. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 2011; 17.
232. FastQC. v0.11.8 ed. Cambridge, England: Babraham Bioinformatics; 2018.
233. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 2012; 1: 18-18.
234. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using De Bruijn graphs. Genome Research 2008; 18: 821-829.
235. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 2012; 19: 455-477.
236. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics 2016; 32: 3380-3387.
237. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 2015; 31: 587-589.
238. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 2011; 27: 578-579.
239. Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics 2012; 13 Suppl 14: S8.
240. Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. Science 1985; 227: 1435.
241. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 2017.
242. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 2015; 13: 278-289.
243. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014; 9: e112963.
244. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 2015; 16: 294.
245. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009; 10: 421.
246. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 2009; 10: R25-R25.
247. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25: 2078-2079.
248. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26: 841-842.

249. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014; 30: 2068-2069.
250. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 2016; 44: D733-745.
251. The Uniprot Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Research 2017; 45: D158-D169.
252. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genomics & Informatics 2009; 23: 205-211.
253. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Research 2003; 31: 371-373.
254. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research 2016; 44: D279-285.
255. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010; 11: 119.
256. Geneious. v8.1.5 build 2015-05-04 17:42 ed. Auckland, New Zealand: Biomatters, Ltd.; 2015.
257. Ncbi Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 2018; 46: D8-D13.
258. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23: 2947-2948.
259. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010; 59: 307-321.
260. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 1985; 22: 160-174.
261. CorelDraw. v19.0.0.0328 x86 ed. Ottawa, Canada: Corel Corporation; 2016.
262. RStudio: Integrated Development for R. v1.2.1335 build1379 ed. Boston, United States of America: RStudio, Inc; 2015.
263. R: A language and environment for statistical computing. v3.6.1 ed. Vienna, Austria: R Foundation for Statistical Computing; 2019.
264. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 2011; 12: 402.
265. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015; 31: 3691-3693.
266. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology 1990; 215: 403-410.
267. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biology 2016; 17: 238.
268. McInerney JO. GCUA: general codon usage analysis. Bioinformatics 1998; 14: 372-373.
269. SumTrees: Phylogenetic tree summarization. V4.0.0 ed: Sukumaran J, Holder MT.; 2015.
270. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. Bioinformatics 2010; 26: 1569-1571.
271. Jackson TA, Saville DJ. Bioassays of replicating bacteria against soil-dwelling pests. In: Bioassays of entomopathogenic microbes and nematodes. Wallingford, Oxon, UK; New York, NY, USA: CABI Pub.; 2000. p. 73-94.
272. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 2016; 44: D733-D745.
273. Hernandez D, François P, Farinelli L, Østerås M, Schrenzel J. De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. Genome Research 2008; 18: 802-809.
274. Schumacher MA, Tonthat NK, Kwong SM, Chinnam NB, Liu MA, Skurray RA, et al. Mechanism of *staphylococcal* multiresistance plasmid replication origin assembly by the RepA protein.

Proceedings of the National Academy of Sciences of the United States of America 2014; 111: 9121-9126.
275. Messer W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. FEMS Microbiological Reviews 2002; 26: 355-374.
276. Datta HJ, Khatri GS, Bastia D. Mechanism of recruitment of DnaB helicase to the replication origin of the plasmid pSC101. Proceedings of the National Academy of Sciences of the United States of America 1999; 96: 73-78.
277. Zhang Z, Nan Z. *Erwinia persicina*, a possible new necrosis and wilt threat to forage or grain legumes production. European Journal of Plant Pathology 2014; 139: 349-358.
278. Rhim, Völksch, Gardan, Paulin, Langlotz, Kim, et al. *Erwinia pyrifoliae*, an *Erwinia* species different from *Erwinia amylovora*, causes a necrotic disease of Asian pear trees. Plant Pathology 1999; 48: 514-520.
279. Kube M, Migdoll AM, Gehring I, Heitmann K, Mayer Y, Kuhl H, et al. Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. BMC Genomics 2010; 11: 393.
280. Mlaga KD, Lotte R, Montaudié H, Rolain JM, Ruimy R. *'Nissabacter archeti'* gen. nov., sp. nov., a new member of Enterobacteriaceae family, isolated from human sample at Archet 2 Hospital, Nice, France. New Microbes and New Infections 2017; 17: 81-83.
281. Brenner DJ, Müller HE, Steigerwalt AG, Whitney AM, apos, Hara CM, et al. Two new *Rahnella* genomospecies that cannot be phenotypically differentiated from *Rahnella aquatilis*. International Journal of Systematic and Evolutionary Microbiology 1998; 48: 141-149.
282. Brady C, Hunter G, Kirk S, Arnold D, Denman S. *Rahnella victoriana* sp. nov., *Rahnella bruchi* sp. nov., *Rahnella woolbedingensis* sp. nov., classification of *Rahnella* genomospecies 2 and 3 as *Rahnella variigena* sp. nov. and *Rahnella inusitata* sp. nov., respectively and emended description of the genus *Rahnella*. Systematic and Applied Microbiology 2014; 37: 545-552.
283. Vadyvaloo V, Martínez-Chavarría L. *Yersinia pestis* and *Yersinia pseudotuberculosis* infection: a regulatory RNA perspective. Frontiers in Microbiology 2015; 6: 956.
284. Robins-Browne RM, Cianciosi S, Bordun AM, Wauters G. Pathogenicity of *Yersinia kristensenii* for mice. Infection and Immunity 1991; 59: 162-167.
285. Souza V, Rocha M, Valera A, Eguiarte LE. Genetic structure of natural populations *Escherichia coli* in wild hosts on different continents. Applied and Environmental Microbiology 1999; 65: 3373.
286. Liu W-Y, Wong C-F, Chung KM-K, Jiang J-W, Leung FC-C. Comparative genome analysis of *Enterobacter cloacae*. PLoS One 2013; 8: e74487-e74487.
287. Townsend SM, Hurrell E, Caubilla-Barron J, Loc-Carrillo C, Forsythe SJ. Characterization of an extended-spectrum beta-lactamase *Enterobacter hormaechei* nosocomial outbreak, and other *Enterobacter hormaechei* misidentified as *Cronobacter* (*Enterobacter*) *sakazakii*. Microbiology 2008; 154: 3659-3667.
288. Wang Q, Yang M, Xiao J, Wu H, Wang X, Lv Y, et al. Genome Sequence of the versatile fish pathogen *Edwardsiella tarda* provides insights into its adaptation to broad host ranges and intracellular niches. PLoS One 2009; 4: e7646.
289. Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschäpe H, Adams LG, et al. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. Infection and Immunity 2002; 70: 2249.
290. Mahlen SD. *Serratia* Infections: from military experiments to current practice. Clinical Microbiology Reviews 2011; 24: 755.
291. Molina-García L, Gasset-Rosa F, Moreno-Del Álamo M, Fernández-Tresguerres ME, Moreno-Díaz de la Espina S, Lurz R, et al. Functional amyloids as inhibitors of plasmid DNA replication. Scientific Reports 2016; 6: 25425-25425.
292. Park K, Han E, Paulsson J, Chattoraj DK. Origin pairing ('handcuffing') as a mode of negative control of P1 plasmid copy number. The EMBO journal 2001; 20: 7323-7332.
293. Chattoraj DK. Control of plasmid DNA replication by iterons: no longer paradoxical. Molecular Microbiology 2000; 37: 467-476.
294. Field CM, Summers DK. Multicopy plasmid stability: Revisiting the dimer catastrophe. Journal of Theoretical Biology 2011; 291: 119-127.

295. Colloms Sean D. The topology of plasmid-monomerizing Xer site-specific recombination. Biochemical Society Transactions 2013; 41: 589.
296. Huber KE, Waldor MK. Filamentous phage integration requires the host recombinases XerC and XerD. Nature 2002; 417: 656-659.
297. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome research 2004; 14: 1188-1190.
298. Kudla B, Caddick MX, Langdon T, Martinez-Rossi NM, Bennett CF, Sibley S, et al. The regulatory gene areA mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. The EMBO journal 1990; 9: 1355-1364.
299. Chen K, Reuter M, Sanghvi B, Roberts GA, Cooper LP, Tilling M, et al. ArdA proteins from different mobile genetic elements can bind to the EcoKI Type I DNA methyltransferase of *E. coli* K12. Biochimica et Biophysica Acta 2014; 1844: 505-511.
300. Taylor BL. Aer on the inside looking out: paradigm for a PAS–HAMP role in sensing oxygen, redox and energy. Molecular Microbiology 2007; 65: 1415-1424.
301. Rebbapragada A, Johnson MS, Harding GP, Zuccarelli AJ, Fletcher HM, Zhulin IB, et al. The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for Escherichia coli behavior. Proceedings of the National Academy of Sciences of the United States of America 1997; 94: 10541-10546.
302. Greer-Phillips SE, Alexandre G, Taylor BL, Zhulin IB. Aer and Tsr guide *Escherichia coli* in spatial gradients of oxidizable substrates. Microbiology 2003; 149: 2661-2667.
303. Horne SM, Mattson KR, Prüß BM. An *Escherichia coli* aer mutant exhibits a reduced ability to colonize the streptomycin-treated mouse large intestine. Antonie van Leeuwenhoek 2009; 95: 149-158.
304. Michel F, Kazuhiko U, Haruo O. Comparative and functional anatomy of group II catalytic introns - a review. Gene 1989; 82: 5-30.
305. Lehmann K, Schmidt U. Group II Introns: Structure and catalytic versatility of large natural ribozymes. Critical Reviews in Biochemistry and Molecular Biology 2003; 38: 249-303.
306. Crozat E, Fournes F, Cornet F, Hallet B, Rousseau P. Resolution of multimeric forms of circular plasmids and chromosomes. Microbiology Spectrum 2014; 2.
307. Stark WM. The serine recombinases. Microbiology Spectrum 2014; 2.
308. Rutherford K, Van Duyne GD. The ins and outs of serine integrase site-specific recombination. Current Opinion in Structural Biology 2014; 24: 125-131.
309. Keenholtz RA, Rowland S-J, Boocock MR, Stark WM, Rice PA. Structural basis for catalytic activation of a serine recombinase. Structure 2011; 19: 799-809.
310. Zellars M, Squires CL. Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. Molecular Microbiology 1999; 32: 1296-1304.
311. Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina TV, Artsimovitch I, et al. Structural basis for transcript elongation control by NusG family universal regulators. Cell 2018; 173: 1650-1662.e1614.
312. Bowen D, Rocheleau TA, Blackburn M, Andreev O, Golubeva E, Bhartia R, et al. Insecticidal toxins from the bacterium *Photorhabdus luminescens*. Science 1998; 280: 2129-2132.
313. Dobrindt U, Hochhut B, Hentschel U, Hacker J. Genomic islands in pathogenic and environmental microorganisms. Nature Reviews Microbiology 2004; 2: 414-424.
314. Busby JN, Landsberg MJ, Simpson RM, Jones SA, Hankamer B, Hurst MRH, et al. Structural analysis of Chi1 Chitinase from Yen-Tc: the multisubunit insecticidal ABC toxin complex of *Yersinia entomophaga*. Journal of Molecular Biology 2012; 415: 359-371.
315. Piper SJ, Brillault L, Rothnagel R, Croll TI, Box JK, Chassagnon I, et al. Cryo-EM structures of the pore-forming A subunit from the *Yersinia entomophaga* ABC toxin. Nature Communications 2019; 10: 1952.
316. Hurst MRH, Jones SA, Binglin T, Harper LA, Jackson TA, Glare TR. The Main Virulence Determinant of *Yersinia entomophaga* MH96 Is a Broad-Host-Range Toxin Complex Active against Insects. Journal of Bacteriology 2011; 193: 1966-1980.

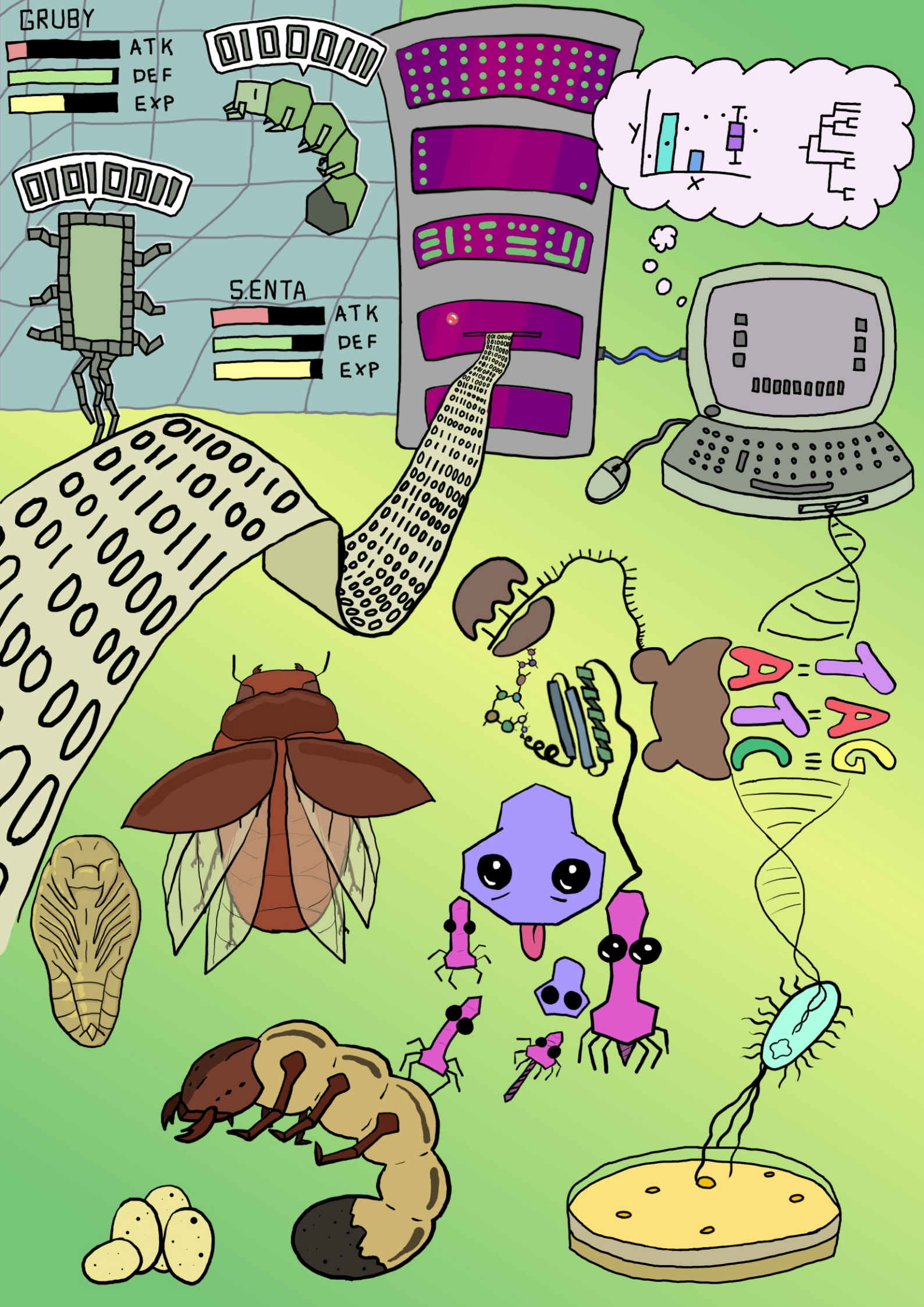
317. Busch A, Phan G, Waksman G. Molecular mechanism of bacterial type 1 and P pili assembly. Philosophical Transactions of The Royal Society A: Mathematical Physical and Engineering Sciences 2015; 373.
318. Godaly G, Svanborg C. Urinary tract infections revisited. Kidney International 2007; 71: 721-723.
319. Krogfelt KA, Bergmans H, Klemm P. Direct evidence that the FimH protein is the mannose-specific adhesin of *Escherichia coli* type 1 fimbriae. Infection and Immunity 1990; 58: 1995-1998.
320. De Masi L, Yue M, Hu C, Rakov AV, Rankin SC, Schifferli DM. Cooperation of adhesin alleles in *Salmonella*-host tropism. mSphere 2017; 2: e00066-00017.
321. Clayton DJ, Bowen AJ, Hulme SD, Buckley AM, Deacon VL, Thomson NR, et al. Analysis of the role of 13 major fimbrial subunits in colonisation of the chicken intestines by *Salmonella enterica* serovar Enteritidis reveals a role for a novel locus. BMC Microbiology 2008; 8: 228-228.
322. Harvey PC, Watson M, Hulme S, Jones MA, Lovell M, Berchieri A, Jr., et al. *Salmonella enterica* serovar Typhimurium colonizing the lumen of the chicken intestine grows slowly and upregulates a unique set of virulence and metabolism genes. Infection and Immunity 2011; 79: 4105-4121.
323. Nair S, Wain J, Connell S, de Pinna E, Peters T. *Salmonella enterica* subspecies II infections in England and Wales – the use of multilocus sequence typing to assist serovar identification. Journal of Medical Microbiology 2014; 63: 831-834.
324. O'Hara CM, Brenner FW, Miller JM. Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. Clinical Microbiology Reviews 2000; 13: 534-546.
325. Kolenda R, Ugorski M, Grzymajlo K. Everything you always wanted to know about *Salmonella* Type 1 fimbriae, but were afraid to ask. Frontiers in Microbiology 2019; 10.
326. Alonso-Caballero A, Schönfelder J, Poly S, Corsetti F, De Sancho D, Artacho E, et al. Mechanical architecture and folding of *E. coli* type 1 pilus domains. Nature Communications 2018; 9: 2758.
327. Hurst MRH, Jackson TA. Use of the green fluorescent protein to monitor the fate of *Serratia entomophila* causing amber disease in the New Zealand grass grub, *Costelytra zealandica*. Journal of Microbiological Methods 2002; 50: 1-8.
328. Bardy SL, Ng SYM, Jarrell KF. Prokaryotic motility structures. Microbiology 2003; 149: 295-304.
329. Wang Q, Suzuki A, Mariconda S, Porwollik S, Harshey RM. Sensing wetness: a new role for the bacterial flagellum. The EMBO Journal 2005; 24: 2034-2042.
330. Jackson TA, Boucias DG, Thaler JO. Pathobiology of amber disease, caused by *Serratia* spp., in the New Zealand grass grub, *Costelytra zealandica*. Journal of Invertebrate Pathology 2001; 78: 232-243.
331. Leader JP, O'Donnell MJ. Transepithelial transport of fluorescent p-glycoprotein and MRP2 substrates by insect Malpighian tubules: confocal microscopic analysis of secreted fluid droplets. Journal of Experimental Biology 2005; 208: 4363-4376.
332. Giddens SR, Tormo A, Mahanty HK. Expression of the antifeeding gene anfA1 in *Serratia entomophila* requires rpoS. Applied and Environmental Microbiology 2000; 66: 1711-1714.
333. Borukhov S, Lee J, Laptenko O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. Molecular Microbiology 2005; 55: 1315-1324.
334. Reddy BL, Saier MH, Jr. Topological and phylogenetic analyses of bacterial holin families and superfamilies. Biochimica et Biophysica Acta 2013; 1828: 2654-2671.
335. Wang I, Smith DL, R. Y. Holins: The protein clocks of bacteriophage infections. Annual Review of Microbiology 2000; 54: 799-825.
336. Schmelcher M, Donovan DM, Loessner MJ. Bacteriophage endolysins as novel antimicrobials. Future Microbiology 2012; 7: 1147-1171.
337. Carolé S, Pichoff S, Bouché J-P. *Escherichia coli* gene ydeA encodes a major facilitator pump which exports l-arabinose and isopropyl-β-d-thiogalactopyranoside. Journal of Bacteriology 1999; 181: 5123.
338. Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics 2012; 13: 162.

339. Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR. Biochemical and genetic analysis of the nifUSVWZM cluster from *Azotobacter vinelandii*. Molecular and General Genetics MGG 1989; 219: 49-57.
340. Freigang J, Diederichs K, Schäfer KP, Welte W, Paul R. Crystal structure of oxidized flavodoxin, an essential protein in *Helicobacter pylori*. Protein Science 2002; 11: 253-261.
341. Higgins CF. ABC transporters: physiology, structure and mechanism--an overview. Research in Microbiology 2001; 152: 205-210.
342. Wilkens S. Structure and mechanism of ABC transporters. F1000prime Reports 2015; 7: 14-14.
343. Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, Kim SH. Crystal structure of the ATP-binding subunit of an ABC transporter. Nature 1998; 396: 703-707.
344. Maróti G, Kondorosi É. Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? Frontiers in Microbiology 2014; 5.
345. Chen F, Guo YB, Wang JH, Li JY, Wang HM. Biological control of grape crown gall by *Rahnella aquatilis* HX2. Plant Disease 2007; 91: 957-963.
346. Łobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnherr H, et al. Genome of bacteriophage P1. Journal of Bacteriology 2004; 186: 7032.
347. Girons IS, Bourhy P, Ottone C, Picardeau M, Yelton D, Hendrix RW, et al. The LE1 bacteriophage replicates as a plasmid within *Leptospira biflexa*: Construction of an *L. biflexa-Escherichia coli* Shuttle Vector. Journal of Bacteriology 2000; 182: 5700.
348. Dente L, Cortese R. pEMBL: A new family of single-stranded plasmids for sequencing DNA. Methods in Enzymology. Methods in Enzymology 1987; 155: 111-119.
349. Qi H, Lu H, Qiu H-J, Petrenko V, Liu A. Phagemid vectors for phage display: Properties, characteristics and construction. Journal of Molecular Biology 2012; 417: 129-143.
350. Betley JN, Frith MC, Graber JH, Choo S, Deshler JO. A Ubiquitous and conserved signal for RNA localization in chordates. Current Biology 2002; 12: 1756-1761.
351. Rousseau P, Gueguen E, Duval-Valentin G, Chandler M. The helix–turn–helix motif of bacterial insertion sequence IS911 transposase is required for DNA binding. Nucleic Acids Research 2004; 32: 1335-1344.
352. Ton-Hoang B, Polard P, Chandler M. Efficient transposition of IS911 circles in vitro. The EMBO journal 1998; 17: 1169-1181.
353. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiology Reviews 2014; 38: 865-891.
354. Guérillot R, Siguier P, Gourbeyre E, Chandler M, Glaser P. The diversity of prokaryotic DDE transposases of the mutator superfamily, insertion specificity, and association with conjugation machineries. Genome Biology and Evolution 2014; 6: 260-272.
355. Ferguson AA, Jiang N. Mutator-like elements with multiple long terminal inverted repeats in plants. Comparative and Functional Genomics 2012; 2012: 14.
356. Mendiola MV, Jubete Y, de la Cruz F. DNA sequence of IS91 and identification of the transposase gene. Journal of Bacteriology 1992; 174: 1345-1351.
357. Garcillán-Barcia MP, de la Cruz F. Distribution of IS91 family insertion sequences in bacterial genomes: evolutionary implications. FEMS Microbiology Ecology 2002; 42: 303-313.
358. Han CG, Shiga Y, Tobe T, Sasakawa C, Ohtsubo E. Structural and functional characterization of IS679 and IS66-family elements. Journal of Bacteriology 2001; 183: 4296-4304.
359. Mahillon J, Chandler M. Insertion sequences. Microbiology and Molecular Biology Reviews 1998; 62: 725-774.
360. Patel M, Raymond B, Bonsall MB, West SA. Crystal toxins and the volunteer's dilemma in bacteria. Journal of Evolutionary Biology 2019; 32: 310-319.
361. Raymond B, Bonsall MB. Cooperation and the evolutionary ecology of bacterial virulence: The *Bacillus cereus* group as a novel study system. BioEssays 2013; 35: 706-716.
362. Kim WS, Park JH, Ren J, Su P, Dunn NW. Survival response and rearrangement of plasmid DNA of *Lactococcus lactis* during long-term starvation. Applied and Environmental Microbiology 2001; 67: 4594.

363. Arturo-Schaan M, Tamanai-Shacoori Z, Cornier M. Stability of plasmid-borne resistance of antibiotics during starvation of *Escherichia coli* in raw and treated waste water and brackish water. Water Science and Technology 1995; 31: 199-202.
364. Tomasz M, Palom Y. The mitomycin bioreductive antitumor agents: Cross-linking and alkylation of DNA as the molecular basis of their activity. Pharmacology & Therapeutics 1997; 76: 73-87.
365. Wei Y, Vollmer AC, LaRossa RA. In vivo titration of mitomycin C action by four *Escherichia coli* genomic regions on multicopy plasmids. Journal of Bacteriology 2001; 183: 2259.
366. Janion C. Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. International Journal of Biological Sciences. 2008; 4: 338-344 pp.
367. Wegrzyn G, Wegrzyn A. Stress responses and replication of plasmids in bacterial cells. Microbial Cell Factories 2002; 1: 2.
368. Silva F, Queiroz JA, Domingues FC. Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. Biotechnology Advances 2012; 30: 691-708.
369. O'Callaghan M, Jackson TA, Glare TR. *Serratia entomophila* bacteriophages: host range determination and preliminary characterization. Canadian Journal of Microbiology 1997; 43: 1069-1073.
370. Loh S, Skurray R, Célérier J, Bagdasarian M, Bailone A, Devoret R. Nucleotide sequence of the psiA (plasmid SOS inhibition) gene located on the leading region of plasmids F and R6-5. Nucleic Acids Research 1990; 18: 4597-4597.
371. Dutreix M, Bäckman A, Célérier J, Bagdasarian MM, Sommer S, Bailone A, et al. Identification of psiB genes of plasmids F and R6-5: Molecular basis for psiB enhanced expression in plasmid R6-5. Nucleic Acids Research 1988; 16: 10669-10679.
372. Ambroise-Thomas P. Emerging parasite zoonoses: the role of host–parasite relationship. International Journal for Parasitology 2000; 30: 1361-1367.
373. Lehnherr H, Yarmolinsky MB. Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 1995; 92: 3274-3277.
374. Christensen-Dalsgaard M, Gerdes K. Two higBA loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilize plasmids. Molecular Microbiology 2006; 62: 397-411.
375. Winther KS, Gerdes K. Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. Proceedings of the National Academy of Sciences of the United States of America 2011; 108: 7403-7407.
376. De la Cruz MA, Zhao W, Farenc C, Gimenez G, Raoult D, Cambillau C, et al. A toxin-antitoxin module of *Salmonella* promotes virulence in mice. PLoS Pathogens 2013; 9: e1003827-e1003827.
377. Wood TL, Wood TK. The HigB/HigA toxin/antitoxin system of *Pseudomonas aeruginosa* influences the virulence factors pyochelin, pyocyanin, and biofilm formation. MicrobiologyOpen 2016; 5: 499-511.
378. Pradel E, Lemaître N, Merchez M, Ricard I, Reboul A, Dewitte A, et al. New insights into how *Yersinia pestis* adapts to its mammalian host during bubonic plague. PLoS Pathogens 2014; 10: e1004029-e1004029.
379. Ghequire MG, Kemland L, De Mot R. Novel immunity proteins associated with colicin M-like bacteriocins exhibit promiscuous protection in *Pseudomonas*. Frontiers in Microbiology 2017; 8: 93.
380. Kim YC, Tarr AW, Penfold CN. Colicin import into *E. coli* cells: A model system for insights into the import mechanisms of bacteriocins. Biochimica et Biophysica Acta 2014; 1843: 1717-1731.
381. Lazdunski CJ, Bouveret E, Rigal A, Journet L, Lloubes R, Benedetti H. Colicin import into *Escherichia coli* cells. Journal of Bacteriology 1998; 180: 4993-5002.
382. Viejo MB, Gargallo D, Ferrer S, Enfedaque J, Regue M. Cloning and DNA sequence analysis of a bacteriocin gene of *Serratia marcescens*. The Journal of General Microbiology 1992; 138: 1737-1743.
383. El Ghachi M, Bouhss A, Barreteau H, Touze T, Auger G, Blanot D, et al. Colicin M exerts its bacteriolytic effect via enzymatic degradation of undecaprenyl phosphate-linked peptidoglycan precursors. Journal of Biological Chemistry 2006; 281: 22761-22772.

384. Braun V. FhuA (TonA), the career of a protein. Journal of Bacteriology 2009; 191: 3431.
385. Pawelek PD, Croteau N, Ng-Thow-Hing C, Khursigara CM, Moiseeva N, Allaire M, et al. Structure of TonB in complex with FhuA, *E. coli* outer membrane receptor. Science 2006; 312: 1399.
386. Patin D, Barreteau H, Auger G, Magnet S, Crouvoisier M, Bouhss A, et al. Colicin M hydrolyses branched lipids II from Gram-positive bacteria. Biochimie 2012; 94: 985-990.
387. Deep A, Kaundal S, Agarwal S, Singh R, Thakur KG. Crystal structure of *Mycobacterium tuberculosis* VapC20 toxin and its interactions with cognate antitoxin, VapB20, suggest a model for toxin–antitoxin assembly. The FEBS Journal 2017; 284: 4066-4082.
388. Winther KS, Gerdes K. Regulation of enteric vapBC transcription: induction by VapC toxin dimer-breaking. Nucleic Acids Research 2012; 40: 4347-4357.
389. Dienemann C, Bøggild A, Winther KS, Gerdes K, Brodersen DE. Crystal structure of the VapBC toxin–antitoxin complex from *Shigella flexneri* reveals a hetero-octameric DNA-binding assembly. Journal of Molecular Biology 2011; 414-540: 713-722.
390. Cruz JW, Sharp JD, Hoffer ED, Maehigashi T, Vvedenskaya IO, Konkimalla A, et al. Growth-regulating *Mycobacterium tuberculosis* VapC-mt4 toxin is an isoacceptor-specific tRNase. Nature Communications 2015; 6: 7480-7480.
391. Matelska D, Steczkiewicz K, Ginalski K. Comprehensive classification of the PIN domain-like superfamily. Nucleic Acids Research 2017; 45: 6995-7020.
392. Gerdes K, Christensen SK, Lobner-Olesen A. Prokaryotic toxin-antitoxin stress response loci. Nature Reviews Microbiology 2005; 3: 371-382.
393. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 2010; 5: e11147.
394. Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, et al. Genetic drift, selection and the evolution of the mutation rate. Nature Reviews Genetics 2016; 17: 704.
395. Ravenhall M, Škunca N, Lassalle F, Dessimoz C. Inferring horizontal gene transfer. PLoS Computational Biology 2015; 11: e1004095.
396. Daubin V, Lerat E, Perrière G. The source of laterally transferred genes in bacterial genomes. Genome Biology 2003; 4: R57-R57.
397. Syvanen M. Horizontal gene transfer: Evidence and possible consequences. Annual Review of Genetics 1994; 28: 237-261.
398. Lawrence JG, Ochman H. Amelioration of bacterial genomes: Rates of change and exchange. Journal of Molecular Evolution 1997; 44: 383-397.
399. Nair RR, Vasse M, Wielgoss S, Sun L, Yu Y-TN, Velicer GJ. Bacterial predator-prey coevolution accelerates genome evolution and selects on virulence-associated prey defences. Nature Communications 2019; 10: 4301.
400. Song H, Gao H, Liu J, Tian P, Nan Z. Comprehensive analysis of correlations among codon usage bias, gene expression, and substitution rate in *Arachis duranensis* and *Arachis ipaënsis* orthologs. Scientific Reports 2017; 7: 14853.
401. Gaur RK. Amino acid frequency distribution among eukaryotic proteins. The IIOABJ Journal 2014; 5: 6-11.
402. Echols N, Harrison P, Balasubramanian S, Luscombe NM, Bertone P, Zhang Z, et al. Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. Nucleic Acids Research 2002; 30: 2515-2523.
403. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. Journal of Molecular Evolution 1986; 24: 28-38.
404. Povolotskaya IS, Kondrashov FA, Ledda A, Vlasov PK. Stop codons in bacteria are not selectively equivalent. Biology Direct 2012; 7: 30-30.
405. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 2013; 30: 772-780.
406. El Karkouri K, Pontarotti P, Raoult D, Fournier P-E. Origin and evolution of *rickettsial* plasmids. PLoS One 2016; 11: e0147492-e0147492.
407. Greve B, Jensen S, Brügger K, Zillig W, Garrett RA. Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. Archaea 2004; 1: 231-239.

408. Ying J, Wu S, Zhang K, Wang Z, Zhu W, Zhu M, et al. Comparative genomics analysis of pKF3-94 in *Klebsiella pneumoniae* reveals plasmid compatibility and horizontal gene transfer. Frontiers in Microbiology 2015; 6.
409. Hingston P, Brenner T, Truelstrup Hansen L, Wang S. Comparative analysis of *Listeria monocytogenes* plasmids and expression levels of plasmid-encoded genes during growth under salt and acid stress conditions. Toxins 2019; 11.
410. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics 1998; 148: 1667.
411. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. Science 2006; 312: 1526-1530.


GRUBY
ATK
DEF
EXP
S.ENTA
ATK
DEF
EXP
y
x
TAG
ATC

Bio-Protection
Bioprotection science for New Zealand

agresearch
āta mātai, mātai whetū

Lincoln University
Te Whare Wānaka o Aoraki

ISBN 978-0-473-54656-4